

## New Process

The present invention relates to new processes for improving the manufacture of clavams e.g. clavulanic acid. The present invention also provides novel DNA sequences and new microorganisms capable of producing increased amounts of clavulanic acid.

Microorganisms, in particular *Streptomyces sp.* produce a number of antibiotics including clavulanic acid and other clavams, cephalosporins, polyketides, cephamycins, tunicamycin, holomycin and penicillins. There is considerable interest in being able to manipulate the absolute and relative amounts of these antibiotics produced by the microorganism and accordingly there have been a large number of studies investigating the metabolic and genetic mechanisms of the biosynthetic pathways (Demain, A.L. (1990) "Biosynthesis and regulation of  $\beta$ -lactam antibiotics." in "50 years of Penicillin applications, history and trends").

*Streptomyces clavuligerus* produces two major groups of antibiotics; one being the cephamycins, cephalosporins and penicillins (Demain, A.L. (1990) *supra*) and the other comprising clavams. Clavams can be arbitrarily divided into two groups, 5S and 5R clavams, dependent on their ring stereochemistry. The commercially important clavam clavulanic acid, a component of the antibiotic Augmentin (trade mark of GlaxoSmithKline), is a 5R clavam. Examples of 5S clavams are clavam-2-carboxylate (C-2-C), 2-hydroxymethyl clavam (2HMC) and alanylclavam (Brown *et al.* (1979) J. Chem. Soc. Chem. pp282-283).

Genes encoding biosynthetic enzymes and regulatory proteins for clavulanic acid production have been located in a cluster next to the genes involved in cephamycin C production and make up a supercluster of antibiotic related genes within the *S. clavuligerus* genome (Alexander *et al.* (1998) J.Bacteriol. 180:4068-79). For example the genes encoding the enzymes involved in clavaminic acid production, a clavulanic acid precursor, which include *orf2* (*ceaS*) (Khaleeli *et al.* (1999) J. Am. Chem. Soc. 121:9223-9224), *orf3* (*bls*) (Bachmann and Townsend (1998) Chem. Commun.:2325-2326), *orf4* (*pah*) (Wu *et al.* (1995) J. Bacteriol. 177:3714-3720), *orf5* (*cas2*) (Marsh *et al.* (1992) Biochemistry. 31:12648-57) and perhaps *orf6* (Kershaw *et al.* (2002) Eur. J. Biochem. 269,2052-2059) are all located within the clavulanic acid cluster. Disruptions in *orfs2-6* cause a complete loss of clavulanic acid production when mutant cultures are grown on starch asparagine medium (Aidoo, K.A. *et al.* (1993) p219-236 In. V.P. Gullo, J.C. Hunter-Cevera, R. Cooper and R. K. Johnson (ed.), Developments in Industrial Microbiology series, vol.33 Society for Industrial Microbiology, Fredericksburg, Va.). However this loss is conditional upon the growth media used for when mutants are grown on Soy medium (Salowe *et al.* (1990) Biochemistry 29: 6499-6508) clavulanic acid production is partially restored (Jensen *et al.* (2002) Antimicrob. Agents and Chemother. 44: 720-726). This phenomenon could suggest that other genes present in the *S.*

*clavuligerus* genome could compensate in some way for the loss of the activity of these genes under certain conditions. Alternatively it could be that the Soy media contains very small amounts one or more of the metabolites produced by the *orfs* 2-6 allowing strains disrupted in these genes to make small amounts of clavulanic acid.

5 Marsh *et al.* (1992) *supra* has reported that *S. clavuligerus* contains two copies of the *cas* gene (*cas1* and *cas2*). *cas1* is not associated with the clavulanic acid gene cluster and has a high homology to *cas2*. Disruption of *cas2* decreases clavulanic acid production by 35% when cultures are grown on Soy medium and eliminates production entirely when cultures are grown on starch asparagine (SA) medium (Paradkar and Jensen 1995 J.Bact 177: 1307-  
10 1314). The disruption of the *cas1* gene results in mutants which produce near wild type levels of clavulanic acid on SA medium, but produce 31-73% less clavulanic acid when grown on Soy medium than the wild type (Mosher et al (1999) Antimicrob. Agents and Chemother. 43: 1215-1224). It is also reported that in mutant strains where both the *cas1* and *cas2* genes have been disrupted no clavulanic acid is produced under any of the fermentation conditions tested.  
15 Interestingly when the genes surrounding *cas1* were sequenced, no additional genes involved in clavulanic acid production were found but instead six novel genes involved in 5S clavam biosynthesis (named *cvm1* to 6) were identified. (Mosher et al (1999) *supra*). Further work on these 5S clavam-specific genes showed that disruption of the genes, using genetic engineering methodologies, leads to improvements in the levels of clavulanic acid made by  
20 the mutant strains and also dramatic reductions in the levels of 5S clavam production (WO98/33896). This reduction in 5S clavam production, in particular the 5S clavam clavam-2-carboxylate, is especially important in the commercial production of clavulanic acid because some 5S clavams are known to be toxic and for this reason the levels are tightly controlled within the British and US Pharmacopoeias.

25 Despite these advances in the understanding of clavulanic acid biosynthesis it is still a highly desirable goal in the pharmaceutical industry to continue to improve production methods for clavulanic acid, both for reasons of cost and for reasons of safety.

The following definitions are provided to facilitate understanding of certain terms used frequently herein:

30 "Gene" as used herein also includes any regulatory region required for gene function or expression.

"*cvm*" genes as used herein refers to any of the genes *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6* or *cvm7* as defined hereinabove.

"*cvmpara*" genes as used herein refers to any of the genes *cvm6para* or *cvm7para* as  
35 defined hereinabove.

"*orf*" genes as used herein refers to any of the genes *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9*, *orf10*, *orf11*, *orf12*, *orf13*, *orf14*, *orf15*, *orf16*, *orf17*, or *orf18* as defined hereinabove.

"*orfpara*" genes as used herein refers to any of the genes *orf2para*, *orf3para*, *orf4para* or *orf6para* as defined hereinabove.

"Disrupted" as used herein means that the activity of the gene (with regard to 5S clavam production) has been reduced or eliminated by, for example, insertional inactivation using an antibiotic resistance gene, preferably apramycin (Paradkar, A.S and Jensen, S.E (1995) *supra*), or other mutagenesis technique (for example those disclosed in Sambrook et al (1989) *supra*). Other mutagenesis techniques include insertion of other DNAs (not antibiotic resistance genes), site-directed mutagenesis to either change one or more bases in the gene sequence or insert one or more bases into the sequence of the gene.

"Deleted" as used herein means that the gene, or a segment thereof, has been deleted (removed) from a larger polynucleotide which, before the deletion was performed, included said gene or segment thereof. When the polynucleotide bearing the deletion is introduced into the genome of the microorganism by means of gene replacement technology (Paradkar and Jensen (1995) *supra*) the activity of the gene or protein encoded thereby is eliminated or reduced such that the levels of 5S clavam produced by the microorganism are reduced. The deletion may be large (for example the complete open reading frame with or without regulatory control regions) or small (for example a single base pair resulting in a frameshift mutation).

"Reduced" as used herein means that the levels of 5S clavam produced by the microorganism of the invention are lower than the levels produced in the corresponding *S. clavuligerus* strain which has not had the relevant open reading frames disrupted or deleted. The corresponding *S. clavuligerus* is therefore the "parent" strain into which the disrupted or deleted open reading frames were subsequently introduced to generate the microorganism of the invention.

"At least maintained" as used herein means that the level of clavulanic acid produced in the microorganism of the invention is the same or greater than that produced in the corresponding *S. clavuligerus* strain which has not had the relevant open reading frames disrupted or deleted. The corresponding *S. clavuligerus* is therefore the "parent" strain into which the disrupted or deleted open reading frames were subsequently introduced to generate the microorganism of the invention.

The present invention concerns new processes for making clavulanic acid using newly identified *S. clavuligerus* genes. Using a probe derived from *orf4* a fragment of the *S. clavuligerus* genome has been isolated and has been shown to comprise a number of genes that when disrupted are shown to affect 5S and 5R clavam biosynthesis in *S. clavuligerus*.

Sequence analysis of the fragment has indicated the presence of a gene showing high similarity to *orf4* (hereinafter called *orf4par*). However surprisingly further sequence analysis of the regions flanking the *orf4par* gene has revealed a new cluster of genes

comprising paralogues of genes previously identified in both the clavulanic acid (*cas2* cluster) and 5S clavam (*cas1* cluster) gene clusters.

Accordingly the invention provides a *S. clavuligerus* microorganism comprising DNA corresponding to one or more open reading frames essential for 5S clavam biosynthesis, wherein said open reading frames are disrupted or deleted such that the production of 5S clavams by said *S. clavuligerus* is reduced and clavulanic acid production is at least maintained, wherein the open reading frames are selected from:

- a) *cvm6para* (SEQ ID NO:1);
- b) *cvm7para* (SEQ ID NO:2);
- c) *cvm6para* and *cvm6* (SEQ ID NO:5); or
- d) *cvm7para* and *cvm7* (SEQ ID NO:6).

In a second aspect the invention provides a *S. clavuligerus* microorganism comprising DNA corresponding to one or more open reading frames essential for 5S clavam biosynthesis, wherein said open reading frames are disrupted or deleted such that the production of 5S clavams by said *S. clavuligerus* is reduced and clavulanic acid production is at least maintained, wherein the open reading frames are selected from:

- a) *cvm6para* and one or more of *cvm1* (SEQ ID NO:7), *cvm2* (SEQ ID NO:8), *cvm3* (SEQ ID NO:9), *cvm4* (SEQ ID NO:10), *cvm5* (SEQ ID NO:11), *cvm6*, *cvm7* or *cvm7para*; or
- b) *cvm7para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm6para*.

The genes *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5* and *cvm6* are disclosed in Mosher et al (1999) *supra* and WO98/33896 (*cvm1* is *orfup1*, *cvm2* is *orfup2*, *cvm3* is *orfup3*, *cvm4* is *ordwn1*, *cvm5* is *orfdown2* and *cvm6* is *orfdown3*). The *cvm7* gene, found to be a further 5S clavam specific gene of the 5S clavam (*cas1*) cluster, has been identified during work leading to the present invention and is disclosed hereinbelow.

In a further aspect the invention provides isolated polynucleotides comprising the *cvm6para* and *cvm7para* open reading frames which are used in the preparation of the *S. clavuligerus* microorganism of the invention. Preferably said polynucleotides comprise open reading frames selected from the group consisting of:

- a) *cvm6para*;
- b) *cvm7para*;
- c) *cvm6para* and *cvm6*;
- d) *cvm7para* and *cvm7*;
- e) *cvm6para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm7para*; or
- f) *cvm7para* and one or more of *cvm1*, *cvm2*, *cvm3*, *cvm4*, *cvm5*, *cvm6*, *cvm7* or *cvm6para*.

In another aspect the present invention provides vectors for cloning and manipulating the *cvm* polynucleotides disclosed herein and which can be used in the preparation of the *S.*

*clavuligerus* microorganism of the invention. Processes for using these vectors to make the *S. clavuligerus* microorganism of the invention are also provided.

The encoded polypeptides from *cvm6para* and *cvm7para* are also provided by the invention (SEQ ID NO:3 and SEQ ID NO:4 respectively).

5 The invention further provides a polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames are selected from the group consisting of:

- a) *orf2para* (SEQ ID NO:12),
- b) *orf3para* (SEQ ID NO:13),
- 10 c) *orf4para* (SEQ ID NO:14), and
- d) *orf6para* (SEQ ID NO:15).

In a further aspect the invention provides a polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames comprise one or more of:

- 15 a) *orf2para*,
- b) *orf3para*,
- c) *orf4para*,
- d) *orf6para*

in combination with one or more genes involved in clavulanic acid biosynthesis selected from  
20 *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9*, *orf10* (Canadian patent application CA2108113 and Jensen, S.E et al (2000) Antimicrob. Agents Chemother 44:720-6) *orf11*, *orf12* (Li, R.N et al (2000) J. Bacteriol 182:4087-95), *orf13*, *orf14*, *orf15*, *orf16*, *orf17*, or *orf18* (patent application PCT/GB02/04989).

Vectors comprising such polynucleotides are also provided by the present invention  
25 together with processes for the use of such vectors to prepare strains of *Streptomyces clavuligerus* which can be used to produce elevated levels of clavulanic acid.

Strains of *Streptomyces clavuligerus* so produced and methods for using them to produce clavulanic acid by fermentation are also provided.

Thus the invention further provides a *Streptomyces clavuligerus* microorganism  
30 comprising a vector comprising a polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames are selected from the group consisting of:

- a) *orf2para*,
- b) *orf3para*,
- 35 c) *orf4para*, and
- d) *orf6para*.

In a further aspect the invention provides a *Streptomyces clavuligerus* microorganism comprising a vector comprising a polynucleotide comprising one or more open reading frames encoding one or more enzymes involved in clavulanic acid biosynthesis wherein said open reading frames are selected from the group consisting of:

- 5 a) *orf2para*,
- b) *orf3para*,
- c) *orf4para*,
- d) *orf6para*

in combination with one or more genes involved in clavulanic acid biosynthesis selected from  
 10 *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9*, *orf10* (Canadian patent application CA2108113 and Jensen, S.E et al (2000) Antimicrob. Agents Chemother 44:720-6) *orf11*, *orf12* (Li, R.N et al (2000) J. Bacteriol 182:4087-95), *orf13*, *orf14*, *orf15*, *orf16*, *orf17*, or *orf18* (patent application PCT/GB02/04989).

The present invention also contemplates a *S. clavuligerus* micororganism comprising a  
 15 combination of one or more disrupted or deleted *cvm6para* or *cvm7para* genes, optionally in combination with other disrupted or deleted 5S genes previously disclosed, together with vectors comprising *orf2para*, *orf3para*, *orf4para* or *orf6para* genes, optionally in combination with other clavulanic acid biosynthetic genes (selected from the genes *orf2* to *orf18*) previously disclosed.

Polynucleotides of the invention can be isolated by conventional cloning methods, such as  
 20 PCR or library screening methods, using the sequences disclosed herein and in Mosher et al (1999) *supra*, WO98/33896, Canadian patent application CA2108113, Jensen, S.E et al (2000) *supra*, Li, R.N et al (2000) *supra* and patent application PCT/GB02/04989, as indicated hereinabove. Examples of such cloning methods are described in, for example, Sambrook, J et al (1989) Molecular cloning, a laboratory manual (2nd Ed) Cold Spring Harbor Laboratory, Cold  
 25 Spring Harbor, New York.

Polynucleotides comprising individual open reading frames can be isolated and ligated together into vectors in a variety of combinations as defined hereinabove using techniques well know in the art. The choice of vector will depend on the function being carried out, for example cloning, expression, gene inactivation or transfer into *S. clavuligerus* eg. for gene replacement. In  
 30 all cases a variety of vectors are available to the skilled person and are well known in the art. For example such vectors are known from Sambrook, J et al (1989) *supra* for general cloning vectors Hopwood, D.A et al (1985) *supra* for Streptomyces vectors, Paradkar and Jensen (1995) *supra*, Mosher et al (1999) *supra* and WO98/33896 *supra* for gene disruption and gene replacement vectors and CA2108113 *supra* for vectors suitable for expression of genes in *Streptomyces*  
 35 *clavuligerus*. However the choice of vector is not limited to just those disclosed in these sources.

Further, in the case of the gene combinations involving the *orf2para*, *orf3para*, *orf4para*, *orf5para* and *orf6para* genes the skilled artisan would be able to design suitable DNA constructs

to ensure that each open reading frame is suitably positioned relative to a transcriptional promoter, whether this be the native promoter or a heterologous promoter that also functions in the *Streptomyces clavuligerus* background, or indeed other regulatory sequence, in such a manner that expression of each open reading frame is optimally achieved.

5 Subsequent manipulation of the polynucleotides, in particular with respect their introduction into the *Streptomyces clavuligerus* background, can be carried out according to standard methods as disclosed in, for example, Hopwood, D.A et al (1985) *supra*. Disruption of gene sequences, and subsequent gene replacement, can be carried out according to the method of Paradkar, A.S and Jensen, S.E (1995) *supra*. Deletion of gene sequences can be carried out using  
10 well established techniques, for example that disclosed in WO98/33896.

Microorganisms of the invention can be prepared from *Streptomyces clavuligerus* strains including, but not limited to, *Streptomyces clavuligerus* ATCC 27064 (American Type Culture Collection, Manassas, Virginia, USA), alternatively available as NRRL 3585 (Northern Regional Research Laboratory, Peoria, Illinois, USA). For example mutant strains of *Streptomyces*  
15 *clavuligerus* can also be used including those prepared by genetic engineering techniques, or those prepared by strain improvement methods. Examples of such strains include *Streptomyces clavuligerus* strains 56-1A, 56-3A, 57-2B, 57-1C, 60-1A, 60-2A, 60-3A, 61-1A, 61-2A, 61-3A or 61-4A as disclosed in WO98/33896.

Thus in another aspect the invention relates to a process for improving clavulanic acid  
20 production in a suitable microorganism comprising isolating a polynucleotide as described hereinabove, manipulating said polynucleotide, introducing the manipulated polynucleotide into a said suitable microorganism, fermenting said suitable microorganism under conditions whereby clavulanic acid is produced, isolating and purifying clavulanic acid so produced. Manipulation of said polynucleotide may be by means of disrupting or deleting gene sequences in the case of  
25 *cvm* genes, optionally together with *cvm* genes, or by inserting into vectors suitable for expression in the case of *orf* genes, optionally together with *orf* genes.

Preferably the suitable microorganism is *Streptomyces clavuligerus*.

Such fermentation, isolation and purification methods are well known in the art, for example the fermentation methods disclosed in UK Patent Specification No. 1,508,977. Methods  
30 for using clavulanic acid in the preparation of antibiotic formulations are similarly well known in the art.

## Examples

### Example 1 - Materials and Methods

35 In the examples all methods are as described in Sambrook, J. et al *supra*, Hopwood, D.A. et al. (1985) *supra* and Kieser, T et al. (2000) Practical *Streptomyces* Genetics, unless

otherwise stated. Transformation methods can also be found in Paradkar, A.S. and Jensen, S.E (1995) *supra*.

### 1.1 Bacterial strains, media and culture conditions.

5        *Streptomyces clavuligerus* NRRL 3585 was obtained from the Northern Regional Research Laboratory (Peoria, IL). *S. clavuligerus* was maintained on either MYM agar (Stuttard, C. (1982) J. Gen. Microbiol. 128:115-121) or ISP Medium #4 agar plates (Difco, Detroit, MI).

10        Cultures for the isolation of chromosomal DNA were grown on a 2:3 mixture of trypticase soy broth and YEME as described by Alexander et al.(1998) J.Bact. 180:4068-79. Cultures for analysis of the production of clavulanic acid and other clavam metabolites were grown on Soy medium (European Patent 0349 121) unless otherwise stated. All liquid cultures were grown at 26°C on a rotary shaker at 250 rpm.

15        Manipulation of DNA in *Escherichia coli* was done using strain XL-1 Blue (Stratagene, La Jolla, CA). *E. coli* cultures were maintained on LB agar medium and grown in liquid culture in LB medium at 37°C (Sambrook, J et al (1989)*supra*). Plasmid-containing cultures were supplemented with appropriate levels of antibiotic.

### 1.2 DNA manipulations.

20        Standard DNA manipulations such as plasmid isolation, restriction endonuclease digestion, generation of blunt-ended fragments, ligation, <sup>32</sup>P labelling of DNA probes by nick translation and *E. coli* transformation were carried out as described in Sambrook J et al (1989) *supra*. Plasmid and genomic DNA isolation from *Streptomyces* spp. was conducted as described in Kieser, T et al (2000) *supra*. Construction of a library of *S. clavuligerus* genomic  
25        DNA fragments in the cosmid pWE15 was carried out according to the manufacturer's instructions (Stratagene).

30        Southern analysis of *S. clavuligerus* DNA fragments was conducted at high stringency as described by Sambrook, J et al (1989) *supra*. Hybridization membranes were washed twice for 30 min at 2xSSC/0.1% SDS and once for 30 min at 0.1xSSC/0.1% SDS, all at 65°C.

## Example 2 - Preparation of the paralogue cluster DNA fragment

### 2.1 Cloning and nucleotide sequencing of the *orf4* paralogue

35        A strong and a very weak hybridization signal was consistently observed on Southern blots of *Nco*I-digested *S. clavuligerus* chromosomal DNA when probed with the *orf4* gene (CA2108113). The strong signal corresponded to the *orf4* gene, but the identity of the gene that gave rise to the very weak signal was unknown. Therefore it was decided to clone this



gene. To this end, *NcoI* fragments from *S. clavuligerus* DNA of approximately 4-5kb in size were ligated into *NcoI* digested pUC120 (Vieira, J and J Messing (1987) Methods Enzymol. 153, 3-11) and screened using a colony blot hybridisation method and employing the *orf4* gene as a probe. Plasmid DNA was isolated from potential positive clones and confirmed to carry a 4.3 kb *NcoI* fragment. A representative clone, pO4H-4, was chosen for further study. The sequencing of the 4.3 kb *NcoI* fragment was carried out. Analysis of the sequence generated identified three genes, one which had homology to *orf4* and was called *orf4par*. The two other genes present were found to have homology with *orf6* and *cvm6* and were therefore called *orf6par* and *cvm6par*. This result suggested that this region of DNA may contain a cluster of genes with paralogues in either the clavulanic acid biosynthetic gene cluster or the *cvm* clavam biosynthetic gene cluster.

## 2.2 Sequencing of DNA flanking the 4.3 kb *NcoI* fragment containing *orf4par*

Sequence analysis of DNA flanking the 4.3 kb *NcoI* fragment containing *orf4par* was achieved by identifying 2 cosmid clones containing the *orf4par* gene. The two cosmid clones containing *orf4par*, 14E10 and 6G9, were isolated from a *S. clavuligerus* pWE15 (Promega, Madison, WI) cosmid bank that had been probed with a 0.46Kb *SalI* fragment that is internal to the *orf4par* gene. These cosmids have been partially mapped using a series of digestions and Southern hybridization experiments (In. Nucleic acid techniques in bacterial systematics. Ed. Stackebrandt, E and Goodfellow, M (1991) John Wiley and Sons, p205-248). Digestion of both cosmids with *EcoRI*, *KpnI* and *NruI* suggest that the insert size of 14E10 is approximately 45 kb and 6G9 is approximately 40 kb. These two cosmid inserts have about 20 kb of overlapping DNA and provided DNA for sequence analysis of regions upstream and downstream of the 4.3 kb *NcoI* fragment containing *orf4par*.

DNA sequence information was generated essentially as described in CA2108113. The DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Pharmacia, Baie d'Urfe, Quebec, Canada) was used. Approximately 13.3 kilobases of contiguous DNA sequence was generated. The nucleotide sequence of the *S. clavuligerus* chromosomal DNA generated in these experiments is shown in SEQ ID No:16.

A number of open reading frames were identified which displayed significant homology with the previously described *orf2*, *orf3*, *orf4*, and *orf6* (CA2108113). These genes have been located within the genome in relation to each other, and are found to be nearly in the same organisation as that of the genes within the clavulanic acid cluster. The genes *orf2par*, *orf3par* and *orf4par* are adjacent to each other and in the same orientation as their counterparts *orf2*, *orf3* and *orf4*. However *cas1* is not downstream of *orf4par* as *cas2* is to *orf4* in the clavulanic acid pathway but is instead within the clavam cluster (Mosher et al (1999) *supra*). Another difference between the clavulanic acid cluster and the paralogue

arrangement is that *orf6par* is end-on-end to *orf4par*, and so is not in the same orientation as *orf2par-4par*, whereas *orf6* is in the same orientation as *orfs2-4* in the clavulanic acid cluster. Surprisingly the gene immediately upstream of *orf6par*, was found to be a gene that had a paralogue in the clavam and not the clavulanic acid cluster. This gene was called *cvm6par*, as it is a paralogue of the *cvm6* gene found clustered with *cas1* (Mosher et al (1999) *supra*). The *cvm6* gene encodes an enzyme that is involved in clavam production (*orfdwn3* in WO98/33896).

Located adjacent to *cvm6par* is a new gene called *cvm7par*. This gene shows homology to *cvm7*, a gene that is located upstream of *cvm3* in the clavam cluster (further described hereinbelow). Upstream of *cvm7* is a new open reading frame, believed to encode a sensor kinase. It encodes a polypeptide of 555 amino acids and shows good similarity to sensor kinase domains of two component response regulator genes.

### 2.3 Functional analysis of the open reading frames

Computer analysis of the DNA sequence shown in SEQ ID No.16 predicts the presence of 7 open reading frames. A description of each gene is shown in Table 1.

**Table 1**

Orf Designation	Homology (blast P)
<i>orf2par</i>	acetolactate synthase (67% identity to <i>orf2</i> carboxyethyl arginine synthase CEAS)
<i>orf3par</i>	asparagine synthetase (49% identity with <i>orf3</i> $\beta$ -lactam synthase BLS)
<i>orf4par</i>	amidinohydrolase (71% identity with <i>orf4</i> amidinohydrolase PAH)
<i>orf6par</i>	ornithine acetyltransferase (47% identity with <i>orf6</i> ornithine acetyl transferase OAT)
<i>cvm6par</i>	aminotransferase (66% identity with <i>cvm6</i> acetylornithine aminotransferase)

<i>cvm7par</i>	Transcriptional regulator (33% identity with <i>cvm7</i> homologue)
Sensor Kinase	Sensor Kinase 47% identity with 2 component system from <i>S.coelicolor</i> A3 (2)

To assess the possible roles of these ORFs in the biosynthesis of clavulanic acid and/or clavams produced by *S. clavuligerus*, insertional inactivation mutants were created by gene replacement essentially as described by Paradkar and Jensen (1995) *supra*.

- 5 However, in order to definitively define the phenotype of these disruptions, it was considered important to disrupt *orf3par*, *orf4par*, *orf6par* and *cvm6par* not only in wild type *S. clavuligerus*, but also in strains of *S. clavuligerus* that were already defective in the expression of *orf3*, *orf4*, *orf6*, and *cvm6* respectively. The *orf3*, *4* and *6* mutants were made as described in United States Patent No. 6,332,106 and the *cvm6* mutant made as described in
- 10 WO98/33896.

### Example 3 – Analysis *orf 4*, and *orf4par*

#### 3.1 Construction of *orf4* mutants

- Mutants disrupted in *orf4* (*pah*) were made as described in United States Patent No.
- 15 6,332,106.

#### 3.2 Construction of *orf4par* mutants

- pO4H-4 (4.3kb *Nco*I fragment cloned into the *Nco*I site of pUC120 (Vieira and Messing 1987 *supra*) was digested with *Kpn*I (one site in the cloned fragment and one site in the
- 20 vector) and religated to reduce the size of the *orf4par*-bearing DNA insert to 1.7kb thereby generating the plasmid p4K-1. The *orf4par* gene within p4K-1 was disrupted by digestion at its centrally located *Eco*NI site and insertion of the apramycin (*apr*) resistance gene cassette from pUC120apr (Trepanier et al. (2002) Microbiology 148: 643-656) after both fragments had been made blunt by treatment with the Klenow fragment of DNA polymerase I. The
- 25 *Kpn*I/*Nco*I insert carrying the disrupted *orf4par* gene was then inserted into the *Eco*RI site of pDA501 after blunting the ends of both insert and vector. pDA501 is a shuttle vector prepared by fusing the *Streptomyces* plasmid pIJ486 (Kieser, T et al (2000) *supra*) to the *E.coli* plasmid pTZ18R (Stratagene) by means of their *Eco*RI and *Bam*HI sites. The resulting construct, 6pDAB, was used to transform *S.lividans* TK24, and finally wild-type *S.*
- 30 *clavuligerus* to thiostrepton (thio at 5µg/ml) and apramycin (apr at 20µg/ml) resistance.

Gene replacement mutants were generated as described by Paradkar and Jensen (1995) *supra*.

### 3.3 Construction of *orf4/orf4par* mutants

5 An approach was undertaken to generate the double mutant by transforming protoplasts of the *orf4par* (*apr*<sup>r</sup>) mutant with the *orf4* (*thio*<sup>r</sup>) disruption construct (Aidoo *et al.* (1994) Gene. 147:41-6). Protoplast preparations from *orf4par* mutants, were transformed with the *orf4* disruption construct isolated from *S.lividans*. Transformants were selected on thiostrepton at 5µg/ml and hygromycin (*hyg*) at 50µg/ml. Primary transformants were put  
10 through two rounds of sporulation under non- selective conditions in order to generate gene replacement mutants as described by Paradkar and Jensen (1995) *supra*.

### 3.4 Fermentation analysis of *orf4*, *orf4par* and *orf4/orf4par* mutants

To test the effect of disrupting *orf4*, *orf4par* and *orf4/orf4par* on clavulanic acid  
15 biosynthesis, spores from each isolate were inoculated into 20ml of seed medium (European patent 0 349 121) and grown for 2 days at 26<sup>0</sup>C with shaking. 1ml of the seed culture was then inoculated into a final stage Soy medium (European Patent 0349 121) and grown at 26<sup>0</sup>C for up to 3 days with shaking. Samples of final stage broth were withdrawn after three days growth and assayed for clavulanic acid productivity by HPLC (Mosher *et al* (1999) *supra*)  
20 and/ or using an imidazole derivatised colorimetric assay (Bird, A.E. *et al* (1982) Analyst, 107: 1241-1245 and Foulston, M. and Reading, C. (1982) Antimicrob. Agents Chemother., 22:753-762).

#### Fermentation analysis of *orf4* disruptant

The *orf4* disruptant was fermented in Soy medium and compared to wild type *S.*  
25 *clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 71%.

From these results it can be concluded that *orf4* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

#### 30 Fermentation analysis of *orf4par* disruptant

Mutant SpDA defective in the *orf4par* gene was fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 12%.

From these results it can be concluded that, like *orf4*, *orf4par* contributes to  
35 clavulanic acid biosynthesis as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

#### Fermentation analysis of *orf4/orf4par* disruptants

When mutants A4-A1 and 3A3-A3, defective in both copies of the *orf4* genes were grown in Soy medium production of clavulanic acid could not be detected.

From these results it can be concluded that under the conditions tested, both genes, *orf4* and *orf4par*, contribute to clavulanic acid biosynthesis as the double disruption, results in  
5 a mutant unable to make clavulanic acid.

### 3.5 Southern Analysis

The *orf4*, *orf4par* and *orf4/4par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copies of the relevant  
10 genes had been disrupted as expected.

### **Example 4 - Analysis of *orf6* and *orf6par***

#### 4.1 Construction of *orf6* mutants

*orf6* mutants were made as described in United States Patent No. 6,332,106  
15

#### 4.2 Construction of *orf6par* mutants

The *orf6par* gene was disrupted by introduction of a neomycin resistance gene (*neo<sup>r</sup>*) into the *RsrII* site, approximately midway through the coding region. In order to achieve this pO4H-4 was digested with *KpnI* to remove *orf4par* and self ligated to give p5K-6. p5K-6 was  
20 digested with *RsrII* and the neomycin resistance gene, released from pFDNeo-S (Denis and Brzezinski (1992) Gene 111:115-118.) as a *PstI/EcoRI* fragment, was inserted after both fragments had been made blunt by treatment with the Klenow fragment of DNA polymerase I. The construct pNeo5K-6A was obtained which has the *neo<sup>R</sup>* gene in the same orientation as the *orf6par* gene.

25 A shuttle vector called pNeo5K-6Atsr#14 was constructed by inserting pIJ486 as a 6.2 Kb fragment linearised with *BglII*, into the *BamHI* polylinker site of pNeo5K-6A. The shuttle vector was used to transform *S. lividans* TK24 and finally *S. clavuligerus* WT to thiostrepton (5µg/ml) and neomycin (50µg/ml) resistance. Primary transformants were subjected to two rounds of sporulation under non- selective conditions in order to generate  
30 gene replacement mutants as described by Paradkar and Jensen (1995) *supra*.

#### 4.3 Construction of *orf6/orf6par* mutants

*orf6/orf6par* double mutants were generated by transforming protoplasts of the *orf6par* (*neo<sup>r</sup>*) mutant with the *orf6(apr<sup>r</sup>)* disruption construct (Mosher et al (1999) *supra*). Protoplast  
35 preparations from *orf6par* mutants, were transformed with the *orf6* disruption construct isolated from *S.lividans*. Transformants were selected on apramycin (*apr*) at 50µg/ml. Primary

transformants were put through two rounds of sporulation under non- selective conditions in order to generate gene replacement mutants as described by Paradkar and Jensen (1995) *supra*.

5     **4.4 Fermentation of *orf6*, *orf6par* and *orf6/orf6par* mutants**

To test the effect of disrupting *orf6*, *orf6par* and *orf6/orf6par* on clavulanic acid biosynthesis, spores from each isolate were tested as previously described in section 3.4.

**Fermentation Analysis of *orf6* mutants**

10     Mutant 6-1A defective in the *orf6* gene was fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 57%. From these results it can be concluded that *orf6* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

**Fermentation Analysis of *orf6par* mutants**

15     Mutant 14-2B(2) defective in the *orf6par* gene was fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 27%. From these results it can be concluded that, like *orf6*, *orf6par* contributes to clavulanic acid biosynthesis as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

20     **Fermentation Analysis of *orf6/orf6par* mutants**

Two separate mutants defective in both *orf6* and *orf6par* were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by an average of 65%.

25     From these results it can be concluded that both *orf6* and *orf6par* are necessary for efficient production of clavulanic acid since disruption of either copy of the gene causes a reduction in clavulanic acid production. Inactivation of both copies of the gene caused a further decrease, but not a complete loss of clavulanic acid producing ability.

**4.5 Southern Analysis**

30     The *orf6*, *orf6par* and *orf6/orf6par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copy of the relevant gene had been disrupted as expected.

**Example 5 – Analysis of *cvm6* and *cvm6par***

35     **5.1 Construction of *cvm6* mutants**

Construction of mutants disrupted in *cvm6* has already been described in WO98/33896 (*cvm6* is *orf6*<sup>dn3</sup>).

### 5.2 Construction of *cvm6par* mutants

A 1.7 Kb *SaII* fragment containing *cvm6par* was released from pO4H-4 and ligated into pUC118 at the *SaII* site. The resulting plasmid was digested with *EcoNI* to release a 140 bp fragment internal to *cvm6par*. In place of this fragment, the neomycin resistance gene from pFDNeo-S, released as an *EcoRI/PstI* fragment, was ligated into *cvm6par* after both fragments had been made blunt by treatment with the Klenow fragment of DNA polymerase I. The neo<sup>R</sup> marker was inserted in the same orientation as *cvm6par*. The neomycin containing *SaII* fragment was released with *EcoRI* and inserted into the shuttle vector pUWL-KS (Weimeier, U.F (1995) Gene 165:149-150.) at the *EcoRI* site. The construct was named pNeoSal1.7U.

The plasmid pNeoSal1.7U was used to transform *S.lividans* TK24, and finally *S.clavuligerus* wild type. The resulting *cvm6par::neo* transformants were selected on MYM medium with 50µg/ml neomycin and 5µg/ml thiostrepton and then subjected to two rounds of sporulation under non- selective conditions to give double cross-over mutants.

### 5.3 Construction of *cvm6/cvm6par* mutants

The construct pNeoSal1.7U isolated from *S.lividans* TK24 was also used to transform the *cvm6* mutant 56-3A, where the apr<sup>R</sup> cassette was inserted into *cvm6* in the same orientation as the gene. Transformants were grown on MYM medium with 50µg/ml neomycin and 5µg/ml thiostrepton. The mutants were put through two rounds of sporulation under non- selective conditions as described above and double cross-over mutants were isolated.

### 5.4 Fermentation of *cvm6*, *cvm6par* and *cvm6/cvm6par* mutants

To test the effect of disrupting *cvm6*, *cvm6par* and *cvm6/cvm6par* on β-lactam biosynthesis, spores from each isolate were tested as previously described in section 3.4.

#### Fermentation Analysis of *cvm6* mutants

It was reported in WO98/33896 that mutants 56-1A, 56-3A, 57-1C and 57-2B defective in the *cvm6* gene produced elevated levels of clavulanic acid (125-141% of the control strain) and greatly reduced levels of clavam-2-carboxylate and 2-hydroxymethylclavam when cultured in Soy medium.

These results suggest that the *cvm6* gene is required for efficient production of the 5S clavams. Disruption of *cvm6* not only results in a reduction in clavams but also a simultaneous increase in clavulanic acid.

#### Fermentation Analysis of *cvm6par* mutants

Mutants 3A1, 3A2, 2A-6, 2B-1 and 2B-2 defective in the *cvm6par* gene were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of  $\beta$ -lactam metabolites. After 72hrs growth, accumulations in clavulanic acid were increased by 6-11%. Production of clavam-2-carboxylate and alanyl clavam was abolished and levels of 2-hydroxymethyl clavam reduced by 50-85%.

These results suggest that like *cvm6* the *cvm6par* gene is required for efficient production of the 5S clavams. Disruption of *cvm6par* not only results in a reduction in clavams but also a simultaneous increase in clavulanic acid.

#### Fermentation Analysis of *cvm6/cvm6par* double mutants

Mutants A-1, A-2, B-1, B-2, C-1 and C-2 defective in both the *cvm6* and *cvm6par* genes were grown in Soy medium and compared to wild type *S. clavuligerus* for their production of  $\beta$ -lactam metabolites. Production of clavulanic acid was increased by 12-27%, production of alanyl clavam and clavam-2-carboxylate eliminated and levels of 2-hydroxymethyl clavam reduced by 70-83%.

These results indicate that, like the *cvm6* and *cvm6par* single mutants, the *cvm6/cvm6par* double mutants produced elevated levels of clavulanic acid and both genes are required for the efficient production of 5S clavams.

#### 5.5 Southern Analysis

The *cvm6*, *cvm6par* and *cvm6/cvm6par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copies of the relevant genes had been disrupted as expected.

### **Example 6 – Analysis of *orf3* and *orf3par***

#### 6.1 Construction of *orf3* mutants

Mutants disrupted in *orf3* were made as described in United States Patent No. 6,332,106.

#### 6.2 Construction of *orf3par* mutants

The plasmid p5.7EcoRI ref (pJOE based hyg) was used as the disruption template for *orf3par*. The insert in this plasmid is approximately 5.7kb and includes part of *cvm6par*, all of *orf6par*, *orf4par*, *orf3par* and part of *orf2par* all carried within the plasmid pJOE829 (Kieser, T et al. (2000); Aidoo et al. (1994) Gene. 147:41-6). The disruption vector was constructed by ligation of a thiostrepton resistance cassette (Aidoo et al. *supra*) into *FseI* digested p5.7EcoRI. A unique *FseI* site is located within the insert 507 bp from the start of *orf3par*.

The correct construct was obtained and used to sequentially transform *S. lividans* TK24 and then *S. clavuligerus* wild type. Primary transformants were selected on thiostrepton (5 $\mu$ g/ml)



and hygromycin (25µg/ml). The mutants were put through two rounds of sporulation under non- selective conditions as described above and putative double cross-over mutants were isolated.

5     6.3 Construction of *orf3/orf3par* mutants

          The *orf3par* disruption cassette described in section 6.2 was isolated from *S.lividans* TK24 and used to transform *orf3::apra* mutants. Transformants were selected on MYM medium containing thiostrepton (5µg/ml) and hygromycin (25µg/ml). The mutants were put through two rounds of sporulation without selection and double crossover mutants isolated as  
10    previously described.

6.4 Fermentation Analysis of *orf3*, *orf3par* and *orf3/orf3par* mutants

          To test the effect of disrupting *orf3*, *orf3par* and *orf3/orf3par* on clavulanic acid biosynthesis, spores from each isolate were tested as previously described in section 3.4.

15    Fermentation Analysis of *orf3* mutants

          Mutants Ap3-1, Ap3-2 and Ap3-3 were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulations in clavulanic acid were reduced by 31-71%.

          From these results it can be concluded that *orf3* is required for efficient production of  
20    clavulanic acid as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation of *orf3par* mutants

          Mutants 3A-1 and 3A-2 were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulations in  
25    clavulanic acid were reduced by 9%.

          From these results it can be concluded that *orf3par* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

Fermentation of *orf3/orf3par* mutants

30    Clavulanic acid biosynthesis was completely abolished when mutants 11-1, 11-2, 2-1 and 2-2 defective in both copies of the *orf3* gene were grown in Soy medium and compared to wild type *S. clavuligerus* .

          These results demonstrate that under the conditions tested, both genes, *orf3* and *orf3par*, contribute to clavulanic acid biosynthesis as the double disruption results in a mutant  
35    unable to make any clavulanic acid.

6.5 Southern Analysis

The *orf3*, *orf3par* and *orf3/orf3par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copies of the relevant genes had been disrupted as expected.

## 5 **Example 7 - Analysis of *orf2* and *orf2par***

### 7.1 Construction of *orf2* mutants

Mutants disrupted in *orf2* were originally made as described in United States Patent No. 6,332,106. These original *orf2* mutants were subjected to a second round of gene replacement to remove the apramycin resistance gene and replace it with a simple frameshift mutation. The plasmid construct used to create the original *orf2* mutant consisted of a 2.1 kb *EcoRI/BglII* fragment of *S. clavuligerus* DNA carried on a pUC119/pIJ486 shuttle vector, with the *orf2* gene disrupted by insertion of an apramycin resistance gene cassette into a centrally located *NotI* site (United States Patent No. 6,332,106). The disruption plasmid construct used in the second round of mutation was derived from the original disruption plasmid by digestion with *NotI* to release the apramycin resistance gene cassette, treatment with the Klenow fragment of DNA polymerase I to fill in the overhanging ends, and then re-ligation to circularize the plasmid. The resulting plasmid construct carries the entire *orf2* gene but with a frameshift introduced at the location of the destroyed *NcoI* site. The construct was used to sequentially transform *S. lividans* TK24 and then the original *S. clavuligerus orf2* mutant. Primary transformants were selected on thiostrepton (5µg/ml) and then subjected to two rounds of sporulation under non-selective conditions. Putative double cross-over mutants were identified based on their loss of apramycin resistance .

### 7.2 Construction of *orf2par* mutants

*orf2par* mutants were generated using a PCR-based targeting kit known as REDIRECT (trade Mark of Plant Bioscience Limited, Norwich, U.K). The plasmids pIJ790 and pIJ773, and the host strain *E. coli* BW25113 were supplied as part of the kit. For this particular application, a pair of oligonucleotide primers, KTA14: 5'-CCATCCCGGCGCCCGTCCGATGCGAAGGAGATCTCCATGATTCCGG-  
GGATCCGTCGACC-3' and  
KTA15: 5'-CGGGGCCGGGCATGGTGAAGTTCGTCCTCCACGGTGGTCATGTAGGC-  
TGGAGCTGCTT-3', designed to disrupt the *orf2par* gene by insertion of an apramycin resistance gene, were synthesized. The *orf2par* disruption cassette was generated by PCR using these two primers with the plasmid pIJ773 as template. PCR conditions used were as described in the user instructions except that no dimethylsulfoxide was used. The *orf2par* disruption cassette was then introduced by electrotransformation into *E. coli* BW25113/pIJ790 which had been previously transformed with the *orf2par* bearing cosmid

14E10 (described hereinabove). Cosmid DNA was isolated from transformants after overnight growth at 37°C to promote loss of the pIJ790 plasmid and analyzed to confirm that the *orf2par* gene had been disrupted. *orf2par* disrupted cosmid DNA was then transferred into wild type *S. clavuligerus* by conjugation. Conjugation was carried out as described by Kieser, T et al (2000) *supra* except that AS-1 medium (Baltz, R. H. Genetic recombination by protoplast fusion in *Streptomyces*. Dev. Ind. Microbiol 21 (1980) 43-54) supplemented with apramycin at 50 µg/ml was used for recovery of transconjugants. Apramycin resistant *S. clavuligerus* transconjugants were subjected to one round of sporulation under non-selective conditions in order to generate gene replacement mutants as described by Paradkar and Jensen (1995) *supra*.

### 7.3 Construction of *orf2/orf2par* mutants

The PCR-based targeting procedure used to generate the *orf2par* mutants (section 7.2) was also used to generate *orf2/orf2par* double mutants. In this case the *orf2par* disrupted cosmid DNA was conjugated into the *orf2* mutants described above (section 7.1) rather than into the wild type strain. Apramycin resistant *S. clavuligerus* transconjugants were subjected to one round of sporulation under non-selective conditions in order to obtain unigenomic mutant spores that had undergone gene replacement as previously described.

### 7.4 Fermentation analysis of *orf2*, *orf2par* and *orf2/orf2par* mutants

To test the effect of disrupting *orf2*, *orf2par* and *orf2/2par* on clavulanic acid biosynthesis, spores from each isolate were tested as previously described in section 3.4.

#### Fermentation Analysis of *orf2* mutants

Mutants defective in the *orf2* gene were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulations in clavulanic acid were reduced by 95-98% (Jensen *et al.* (2000) *supra*).

From these results it can be concluded that *orf2* is required for efficient production of clavulanic acid as elimination of this gene by disruption causes a severe reduction in clavulanic acid production.

#### Fermentation analysis of *orf2par* disruptant

Mutants defective in the *orf2par* gene were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, accumulation of clavulanic acid was reduced by 10-30%.

From these results it can be concluded that, like *orf2*, *orf2par* contributes to clavulanic acid biosynthesis as elimination of this gene by disruption causes a reduction in clavulanic acid levels.

#### Fermentation analysis of *orf2/orf2par* disruptants

Mutants defective in both *orf2* and *orf2par* were fermented in Soy medium and compared to wild type *S. clavuligerus* for production of clavulanic acid. After 72hrs growth, no clavulanic acid production could be detected from the strains contain the *orf2* and *orf2par* mutations. These results demonstrate that under the conditions tested, both genes, *orf2* and *orf2par*, contribute to clavulanic acid biosynthesis as the double disruption results in a mutant unable to make clavulanic acid.

#### 5. Southern Analysis

The *orf2*, *orf2par* and *orf2/2par* mutants were further characterised by Southern analysis. The results confirmed that in these mutants the chromosomal copies of the relevant genes had been disrupted as expected.

#### Example 8 –Analysis of *cvm7* and *cvm7par*

Sequence analysis had identified two additional genes in the paralogue cluster that did not have obvious paralogues in either the clavulanic acid or *cvm* gene clusters. It was of interest to determine if either of these genes was a paralogue to an as yet unidentified *cvm* gene. Therefore the sequence of the *cvm* cluster (WO98/33896) was extended downstream of *cvm3* (*orfup3* in WO98/33896).

#### 8.1 Extension of *cvm* cluster sequence

The cosmid 10D7 (described in WO98/33896) was digested with the restriction endonuclease *SacI*. From this digestion a 6.8 kilobase DNA fragment containing *cas1* and *cvm1* was isolated and cloned into a pUC119 based plasmid. The resultant plasmid pCEC019 was used as a template to generate sequence information which allowed completion of the partial *cvm3* gene reported in WO98/33896. In addition, the sequence information showed the presence of another open reading frame, *cvm7*, which was incomplete in this fragment. In order to complete the *cvm7* gene sequence, the next adjacent *SacI* fragment from cosmid 10D7, a 1.9 kb fragment, was subcloned. Sequence information was obtained from the end of the clone which contained the remainder of the *cvm7* gene, up to the point where the start codon for the *cvm7* gene could be identified. In total, this resulted in the generation of a further approximately 3.9 kb of new DNA sequence which is described in Sequence ID No.17.

#### 8.2 Sequence analysis

The size of *cvm7* and its orientation relative to the rest of the *cvm* cluster is showed diagrammatically in fig2. Sequence homology searches demonstrated that this gene shares homology with transcriptional regulator genes. In addition *cvm7* also shared 33% identity

with one of the two genes identified in the paralogue cluster that did not have any obvious  
 paralogues within the known clavulanic acid or clavam biosynthetic genes. Therefore since  
*cvm6* and *cvm6par* have been shown to be paralogues, from this sequence data it can be  
 concluded that *cvm7* and *cvm7par* are paralogues of genes involved in 5S clavam

5 biosynthesis.

#### Brief description of the figures

Figure 1. Diagram of the paralogue cluster. The orientation of transcription is shown for each  
 gene (direction of arrow)

10 Figure 2. Orientation of *cvm7* in relation to published *cvm* cluster (WO98/33896).

Figure 3. Annotated sequence of the paralogue cluster

#### Brief description of the sequences

SEQ ID NO:1 *cvm6para* open reading frame  
 15 SEQ ID NO:2 *cvm7para* open reading fame  
 SEQ ID NO:3 *cvm6para* polypeptide  
 SEQ ID NO:4 *cvm7para* polypeptide  
 SEQ ID NO:5 *cvm6* open reading frame  
 SEQ ID NO:6 *cvm7* open reading frame  
 20 SEQ ID NO:7 *cvm1* open reading frame  
 SEQ ID NO:8 *cvm2* open reading frame  
 SEQ ID NO:9 *cvm3* open reading frame  
 SEQ ID NO:10 *cvm4* open reading frame  
 SEQ ID NO:11 *cvm5* open reading frame  
 25 SEQ ID NO:12 *orf2para* open reading frame  
 SEQ ID NO:13 *orf3para* open reading frame  
 SEQ ID NO:14 *orf4para* open reading frame  
 SEQ ID NO:15 *orf6para* open reading frame  
 SEQ ID NO:16 paralogue cluster  
 30 SEQ ID NO:17 extended *cvm* cluster (underlined sequence denotes new sequence over that  
 disclosed in WO98/33896  
 SEQ ID NO:18 *orf2para* open reading frame (reverse complement)  
 SEQ ID NO:19 *orf3para* open reading frame (reverse complement)  
 SEQ ID NO:20 *orf4para* open reading frame (reverse complement)  
 35 SEQ ID NO:21 *cvm6* polypeptide  
 SEQ ID NO:22 *cvm3* polypeptide  
 SEQ ID NO:23 *orf6para* polypeptide

SEQ ID NO:24 orf4para polypeptide

SEQ ID NO:25 orf3para polypeptide

SEQ ID NO:26 orf2para polypeptide

SEO ID NO:1 cvm6para

5  
10  
15  
20

SEO ID NO:2 cym7para

25  
30  
35  
40  
45  
50

SEO ID NO:3 cvm6para polypeptide

55

60

SEQ ID NO:4 cvm7para polypeptide

65

AGAGAAPASASGSVSASVSGSGSGSAPASVPTFFPGSVSGSASVAASVAAPVSGHVS GPGSAFGSVALHRPQTLRGEF  
 VHGAQGMRTGQVFPTLPFFVGRGDELRLLESATSAFHTSGRVAFVVGEGSGKTRLLSELSRVPDSVRTVWASCSES  
 EDRPDYWPWTVLRHLYAMWPERMHGFPGLRRALAE LPEVGPPEQPHSPDGGEENSNGNDGAGDGDSTPAHTLT LAP  
 5 ALAPPRSREAREFTLHDAVCQALLRTRVREPVVIMLEDMERADAPSLALLRLIVEQLRTVPLLLVVTRTRFRLAHDAELRRA  
 AAVILQSTGARRVLLNALDARATGELAGGMLGKAPDTLLVRALHERSAGNPYFLVQLRLSLRQGLAAAWETEIPDELAGV  
 VLQRLSSVPPAVRRVLDICAVVERSCE RRVETVLRHEGIPLENVRTAVRGGLLEEDPDDPGR LRFVHPLVREAVWDDLE  
 NTRRPVSRSSALGALATV

## SEQ ID NO:5 cvm6

10 GTGCCCCGCTCCGACTCGAAGCACTGGACCGTGCCACCCCTCATCCACCCACCCCTCTCCGGAAACACCGCGGAACGGAT  
 CGTGCTGACCTCGGGGTCCGGCAGCCGGGTCCGCGACACCGACGGCCGGGAGTACCTGGACGCGAGCGCCGCTCCTCGGGG  
 TGACCCAGGTGGGCCACGGCCGGGCGGAGCTGGCCCGGTGCGGCGGAGCAGATGGCCCGGCTGGAGTACTTCCACACC  
 TGGGGACGATCAGCAACGACCGGGCGGTGGAGCTGGCGCACGGCTGGTGGGGCTGAGCCCGGAGCCGCTGACCCGCGT  
 CTACTTCACCAGCGCGGGCGGAGGGCAACGAGATCGCCCTGCGGATGGCCCGGCTCTACCACCACCGGCGCGGGGAGT  
 15 CCGCCCGTACCTGGATACTCTCCCGCCGCTCGGCCCTACCACGGCGTGGGATACGGCAGCGGCGGCGTACCGGCTTCCCC  
 GCCTACCACCAGGGCTTCGGCCCCCTCCCTCCCGGACGTCGACTTCTGACCCCGCGCAGCCCTACCGCCGGGAGCTGTT  
 CGCCGGTTCCGACGTCACCGACTTCTGCTCGCCGAACCTGCGCGAGACCATCGACCGGATCGGCCCGGAGCGGATCGCGG  
 CGATGATCGGCGAGCCGATCATGGCGCGGTGCGCGCCGCGGCCCGGCGACTACTGGCCCCGGGTGCGCCGAGCTG  
 CTGCACTCCTACCGCATCCTGCTGATCTCCGACGAGGTGATCACGGGGTACGGGCGCACCGGGCACTGGTTCCGCCCGCA  
 20 CCACTTCGGCGTGGTCCCGGACATCATGGTCACCGCCAAGGGCATTACCTCGGGGTATGTGCCGACGGCGCGCTCCTG  
 ACCACCGAGGCGCTCGCCGACGAGGTGCTCGGCGACAGGGCTTCCCGGCGGGCTTACCTACAGCGGCCATGCCACGGC  
 CTGCGCGGTGGCCCTGGCCAACCTGGACATCATCGAGCGCGAGAATCTGCTCGACAACGCCAGCACCGTCTGGCGCCTACC  
 TGGGCAACCGCTGGCCGAGCTGAGCGATCTGCCGATCTGCGGAGCTGCGGCGAGACCGGCTGATGCTCGGTGTCGAA  
 CTGGTCGCGGACCGCAACCGGGAGCCGCTGCGGGCGCGCCGCTGCGCGAGGCCCTGCGCGAGCGGGCGGGCATCCT  
 25 GCTGCGCGCAACGGCAACGCCCTCATCGTCAACCCCCCGCTGATCTTACCCAGGAAGACGCCGACGAACTCGTGGCGG  
 GCCTGCGCTCCGTACTCGCCCGACAGGCCGACGGCCGGGTGCTCTGA

## SEQ ID NO:6 cvm7

30 ATGAAGTACGACATAACCCACCATCCGGCCTTCGGTTCGACCTCTCGGCCCGTTGACCGTGACCGCGGGCGAGCAACC  
 CGTGGACCTGGGCGCGCCACGGCAGCGCCCTGCTCGCCCTGCTGCTCATCGATGTGCGCAACGTGGTCCCGCTGCCG  
 TCATGACCGCGCTCGATCTGGGGGGCGACCCACCGTCCCGGCTCGGGGGACGCTCCAGGCTTATGTCTCCGACTGCGG  
 AAACCTCCTGCACCGCCATGACCGTTCCTTCGCCTTGTCACACGCTCCAGGGGTATCTCCTCGAAGTGGATTGGCGAA  
 GGTGGACGCGTGGTTTTTCGAGACACGTGTAGGGAGTGCCGGGAATTGAGCAGGGCCCGGAACCCGAGGCCACCCGGG  
 35 CCGTGGCCTGGTCCGCCCTGGAGATGTGAAGGGCACACCATATGGGCGAGCTGCATGATTATGAATTTGTGGCGGGGAG  
 CGCGACCGGCTGGAAGGAATCCGGTTACGCGCGCTGGAGACCTGGTCCAGGCGTGTCTCGATCTCCAGCACTATGAAGA  
 GGTTCATTTAGCTCGCGGAGGAGATCCACCGCAATCCGGAACCTGGAACGGCTGGGCGGTCTCTTCATGCGGCCGAGT  
 ATCATTCGCGACGCTCGCGGAAGCCCTGTTGACGTATGAACGTATGCGTACCGCGGTGGCGGAGAATCTGGGGCCGAT  
 ATCAGTCCGGAGCTCCAGGAACCTCATGAAAGATTCTGCGCCAGGAACCTCACGAGACACCCGCCGCGCATCGACGGC  
 40 CTCCCTCACACGGGCGCGGGCCCGCACGGGCCCCCGCCCTGGCCGAAACCGGCACCCCGCCGACCCGCGGACATGG  
 CCGAAACACGGTGGCGGAGGAAGCGCCGCGCCCCCGCCCGCGGCGCGCCGGAGACCCCGCCCGCCCATGCCGTCCCC  
 GTACCGCTCCCCCATCCGTACGGGGCGTCCCGCGGTACCCCGGTGCTTCCCGGTCCCCCGTCCGCCCTCCGTTT  
 AGCGGCACCCCGATCCGAGACCGGACCCGCGCGCCCGCTCCCCCTCCGCGGGCGGCGGACTCATCGGCCGCC  
 GCGCCGAACCTGCGCAGGTGCGGCTGCTGCTGACGAAGACCCGCGCGGGCCACGGCCATGTCTGCTGGTCTGCGGCGAA  
 45 CAGGGCATCGGGAAGACCCGGCTCCTGGAGCACACCGAGCACACCTGGCCGCGGGCGGTTCCGGGTGGTCCGTTTCGCA  
 CTGCGTCCGCCACCTCCCGGCACCGGGCTACTGGCCCTGGGAGCACCTCGTACGCCAGCTCGACCCGGACAGCGGCCTCG  
 GTGACGACGGCGACCGGACCCCGTCCGCCAGGCGGATGGCTGCCGAACACCACTACCCACAGATGCGGATCTGC  
 CGGACGGTGTCTCGCGCGCGCGGCGGACCCCGCTCTGTTGATCTTGGAGGATCTGCACCTCGCCACCGCGCGGTCT  
 50 GGATGTGCTCCAGCTCCTGGTCAAACAGATCGGCCAGGCCCGCTCATGGTCTGCGCACCCCTGCGCGAGCACGATCTCG  
 CCCGGGACCCCGCGTCCGCGGGCGGTGGGCGCATCTCCAGCGGGCAACACCGGCACCCCTCCGGTGGACGGGCTC  
 ACCGAGGAGCAGAGCCGGGAGCTGATCTGCTCGGTGCGGGGGCCCCGTTCCGCGCCCATGACGCCAACGGCTCCAGCG  
 CGCCTCGGGCGGCAACCCGTTTCTGCTGCTCAGCATGGTACAGGGGAGGACGGCACCCAGGAGTGGGACGGCCGTGCG  
 55 TCCCGTTCCGAGGTGCGCGAGGTGCTGCACGAGCGGCTGAGCGAATGCTCCCGTCCACCCAGGACGTGCTACGCTCTGC  
 GCCGTGCTCGGCATGAGCGTGCGCCGACCGTGCTCACCGACATCATGTCCACGCTCGACATCCCGCACACCGCGCTCGA  
 CGACGCGCTCGGCACGGGCTGCTGCGCCACGACCGGAACACCGACGGAATGGTCCACTTCGCCCATGGGTGACCCGGG  
 ACTTCCTGCTCGACGACACCCCGCGGTACCCCGCGCCGCTGGCACACCGGGTCCCGCCACCCCTCGCCCTGCGCTTC  
 CAGCAGGGCGACGACACCGCGAGATCCGCCCGCACTGTCTGGCCGCGGCCGCTGCTCGGCGCCCGCGGGGGTGGC  
 60 CCCCCTGCTGGCGCTGGCCGACCGGGAGCAGTCCCGCTTCTCCACGCGGAGGCGCTGCGCTGGAGAGCGGGTTCG  
 CGGTCTGCGCGCGCTGCCCGGGACACCGCGTTCGCCGCTCGAACTCCAGTTGCGCAAACGGATGATGGCGCTGCAC  
 GCGCTGATGGACGGCTATGGATCGGCCGCGTTCGAGACGTTCTCTCCAGGTACCCAGTGGGAACACGCTCTCGACAA  
 CACCCAGCCACCGGGTGTGTCACGTCCAGGCGCTGAGCGCGTACCCAGGGCGCCATGAGCAGGGCGGAGCTGG  
 65 CCGGGTGTGTCACGAGCTGGCCGACACGGCGCGGACCGGAGGCGCGTGGCGGCGCTGCTATGTGGACGGCTCACC  
 CTGATGTGGGCGGACGGTTCGACGAAGCCCTCGCCGCTCGCCAGGGACCGGATCAGGCGCCCTGCTGGCCG  
 ACACCGCAGGACCGCCCGCACGGCGCGGCTCCAGGACCGGCGTATCGACTTCGCGCCTATCTGGCGCTCG  
 GCCACTGTCTCAGCGGCGACCGGATTACAGCCAGCGCTACCGACGGAACCTCTCCACCTACCCAGTCCGAACGGTAC  
 GACCGCGCGTGGGACCGGGCTTCGCCCGCTATGTGGACGCGCTATCGCCGTACGAGAGTGGATGTCCAGGGGGTGTG  
 GCTGGCCGCGCGGGCGGGGCTCGACCTCGCCGCGCGTGCAGCTCCCGTCTGGCAGCGGATGCTCGCCGTCCCCCTCG  
 GCTGGGCCGAGGTCCACAGGGGGCGACACAAGGGGCTGGCCGGATGCGGGAGGCGCTGCACGAGGCGGCCCGGCAC



CGGACCCTGCTGCGCCGTACGCTCCACCTCGGCCTGCTCGCCGACGCCCTCCAGTACACGGGCGCCCGGGAACAGGCCCG  
GCGACGATGTCCTCCGCCGTACGGGAGATCGAGCGCCGCGGCGAGTACTTCTGTCTCCGGCCGAGTGGCCCTGGGCC  
GGCTCCTCCACAGCCACGGCACCTCCGCCGCGGCGGAGCACCGGGTCTGCCACGGCAGGCACTGA

## 5 SEQ ID NO:7 cvm1

ATGTCCCGCTCTCCGCCCGAGTCCCCGCCGGTTCCGTGTCCGCCGCGGTTCCGCGTCCGCCGGTCCGCGCCCTGCGGGA  
CCTTCCGGTCAGTGCCAGGGGCTCGGCTGCCTGCCGACCACCGACTTCTACGGACGCCCGGACCGCGCCCGGGCAGCG  
CCACCATCCGCGCCCGCTCGACGCCGGGTCAACCTGTGGACACCGCCGACGTCCAGGGGCTCGGCGCCGGTGAGGAG  
CTGCTCGGACGGGCGGTGCGGGGCCGCGGGACGAGGTCTGATCGCCACCAAGTTGGCATGGTGCCTCGTCCGACGG  
10 CGCCTCCAGGGCTTGTGCGGCGAGCCGTCTACGTCCGCGCGGCTGCGAACGGTCCCTGCGTCTCGGCACCGACC  
GCATCGACCTGTACTACCAGCACTGGACGGACCCGGCGGTGCCGATCGAGGAGACCGTGGGTGCGGTGGCCGAGCTGGTG  
CGCGAGGGCAAGGTCCGACAGGCTCGGTCTCTCGAGCCCTCCGCGGCCACGCTGCGCCGGGCGGACCGGCTGCACCCGGT  
GACGGCGGTGACAGCGAGTGGAGCCTGTGGTCCGCGGGATCGAGGACGAGGTGGTGCCTGTGCGGGAGCTGGGGA  
TCGGGATCGTCTGCTTACGCCCTCTGGGACGGGGTTTTCTCACCGGCACCATCCGCACACCGACGATCTGGGGACGAG  
15 GACTTCCGCCGGGGCCAGCCCCGGTTACGCGCTCCGGCCCTCGCGCGCAACCGCTCGTTGCTGCACCGGCTGCGCCCGGT  
CGCGGACGGTCTGGGGCTGACCCTGGCACAGCTCGCGCTCGCCTGGCTGCACCACCGGGCGAGGACGTCGTCCCGATCC  
CGGGCACCGCAACCCGGCCCATCTCGCGGACAATCTCGCCGCCCTCGATCCGGTGGACGACCGGTCCCTCGCGGAG  
GTGACGGCCGCGATCTCCACCCGGTGTCCGGGGAGCGGTACACCCGGCATTGCTCGCCATGATCGGCAACTGA

## 20 SEQ ID NO:8 cvm2

ATGTCCGTGGCATCGGCCGTATGACGGACGAGCAGCGCAAGGCGGTATCACCGCTACTTCAAGGCGTTGACAAACGG  
CGGCGTCCGACGACGGCACCCCCGCGATCGACTACTTCGCCGAGGACGCGGTCTTCTTCTTCCCAAGTGGGGTCTGG  
CCCGGGGCAAGTCCGAGATCGCCCGGCTCTTCGACGACCTCGGGGGCACCATCCGCTCGATCACCCACCATCTGTGGTCC  
GTCAACTGGATTCTGACCGGGACCGAACTCTCGCCGCGGAGGGCACCACCACGGTGAGCACCGGGACGGGCGGTGGCG  
25 GCGGGGTGACCCGAGTGGGCGCGCGGCGGTGGTGACGGTCTACGAGGTGCGGGACTTCTCGTCCACCGGGCCTTCG  
TCTATCTGGACCCCGATTACGCGGGCAAGGACACCGCGCTTACCCGTGGCTGTGA

## SEQ ID NO:9 cvm3

GTGACCCGGCTCCGGGCCTTTCGCGCACACCCACGGGTCCGTGTCCGGGAGTCTGCTGCGCCGGGTGGCGGGCCACTA  
30 TCCACCGGGGTGGTCTTGGTCAACCGTCCGGCCGAGGCTCCGGGGCAGCCGCCGCCCGCCATGGTGGTGGGGACGTTCA  
CCTCGGTGTGCTCGATCCGGTGTGTTGGGTTTCTCCCGCCAGGTCTGTCGACGACCTGGCCGCGGCTCCGGGCGGCC  
GGGCGTTTCTGCGTCAATGTGCTCGGCGCGGATCAGGGCCCGGTCTGCGGAGTTTCGCCGGGGCGATCCGGGGCGCTG  
GGAGGTGCGGTACCGGACGACGGCCACCGGCTCCCCGCTCTGCTCGACGCGCTCGCGTGGTTCGACTGCGAGGTGGCGG  
GGGAGACGGAGGCGGGCGACCACTGGTTCGTACCGGGGCGGTGCGCGACCTCGGGGTGATCCGCGAGGGTTGCCCCCTG  
35 GTCTTCTGCGGGGCGACTACGGGCACCTGGGCCGGGGCGCGGCTCGGGCCGGGCGGGGCGGCGGTCCGCCGTCTGCC  
GGTCTGA

## SEQ ID NO:10 cvm4

GTGGAATGCCGCATATTGAGATCGACGAACTGCCGTTGCTGGACGGGGAGGTCTCGGGGACGCCCGGATCGGTTACGC  
40 CATGTACGGCACGCCGAACGCCGACGGGACGAACGTGGTGTCTGTCCGTCTTCTCGGCCGGGACCACACCGGGTACG  
ACTGGCTGATCGGTGCGGGGCTGCCGCTGGACACCCGGCGGTACTGCGTCTGTCACCGCCGGACTCTTCGCAACGGGGT  
TCCAGCTCGCCCGGAACCAACCCGTCCGGGTCCCGCTTTCGCGTGTACTCCGACAGCAATGTGCGGCGCAGCACCG  
GCTGCTCACCGAGGAGCTGGGGTACGGGAACCTGGCCCTGGTACGGGCTGGTTCGATGGGCGCGGGCCACGCCTACAGT  
GGCCGTGTGCTGCTACCGGGGATGGTGCGCGCGGATCGCCCGCATCTGCGGGGCGCGGCTGAGCGGGCGGACGAGGCTG  
45 CTGCTGTCCGGTCTGGCCGCGGCGCTCAGCGCCGACCGCGGGGAGCGGGGGCGGAAGGCGGCGGGCGGGTGTTCGCGG  
GTGGGGACCTCGCGTTCTTCTGGGCCCCCGTGCCACCGGGAGCTGGGTTTCGCCACCCGCGAGGAGTACCTACCG  
GCTTCTGGGAGCAGGTCTTCTCTCGGGCCCGGCGCGCGGATCTGCTCACCATGGTGCACCTGGGAGAACACGGAT  
GTGGGGGCGACACCCGGGGCGGGGGGAGCGTGCAGGCGGCGCTGGCCTCCGTACGGCGCGGGCCGTGGTGTGCGGG  
CGCCCTGGACGTGTGTTTCGCGCTGAGGACGAGAAGCGGTGGCCGATCTGCTGCCGTATGCTCGTGGAGGTGATCC  
50 CGGGAGTGTGGGGGCATCTCGCGGGGTCCGGGGGTCCGCCGCGGACCGGGAGTTCATCGGGGGCGCGCTGCGGGCGGCTG  
CTGGACAGCCCGGTGGACGGGGGCTGA

## SEQ ID NO:11 cvm5

GTGAAGTCCATTCTCTTCTATCTGCCAACGGTCCGGCAGTCATGCGCAGGTCCAGCGGGGTATGGCGGGGGTCAATCCGCA  
55 GAACTACCAGAACATGCTCCGGCAGCTCACCCGGCAGGCGCAGGCGGCCGACGAACCTCGGCTACTGGGAGTGTCTTCA  
CCGAGCACCACTTCCACACCGAGGGTTTCGAGGTCTCCAACAACCCGATCATGCTGGGGCTCTACCTCGGCATGCAGACC  
CGGCACATCCGGGTCCGCCAGATGGCCAACGTCTGCGCTGCACAATCCGCTGCGGCTGGCCGAGGATCTGGCGATGCT  
CGACCACATGACCCGGGGCCGCGCCTTCGTCCGGATCGCGCGCGGGTTCAGAAGCGCTGGGCCGACATCATGGGCGAGG  
TGTACGGGGTCCGCGGCACCTGTCCGACGCGGGGAGCGGGACCGCGCAATCGTGCCCTCTTCGAGGAGCACTGGGAG  
60 ATCATCAAGAAGGCGTGGACGACCGAGACGTTCAACCACTCCGGGGAGCAGTGGACGATCCCGGTGCCGACCTGGAGTT  
CCCCTACGAGGCTGCGCGCTACGGCCGGGGCTCGACGAGAACGGCGTCATCCGCGAGGTGGGCATCGCGCCCAAGC  
CCTACCAGCGCCCCACCGCCCGTCTTCCAGCCGTTCACTTCAGTGAGGACACGTTCCGGTTCTGTGCCCGGGAGGGC  
GTGGTGCCGATCTGATGAACACCGACGACAGATCGTCGCCCGGTGATGGACATCTACCGGGAGGAGGCCGAGGCGGC  
GGGCCACGGCACCTGCGGGCGGGCGAGCGGGTCCGGGTGATGAAGGACGTCCTGGTCTCCCGGGACTCCGGCGAGGCC  
65 ACCACTGGGCGTCCCGCGGGCGGGGCTTCATCTTCGAGAAGTGGTTCGGCCCCATGGGCTTACCGAGGCGCTGCGCGCG  
ACCGGCGAGACGGGTCCGATCGGCTCGGACTACAAGACCCTGGTCGACCGGGGGCTGGAGTGGGTCCGCACCCCGGACGA

CATCAACCGCATGATCGAGAAGCTGGTGGAGCGGCACGATCCGGAGTATCTGCTCCAGTGCCAGTACTCCGGGCTGATCC  
CGCACGATGTCCAGCTGCGCAGCCTGGAGCTGTGGGCCACCGAGATCGCCCCCACTGGCTCTGA

## SEQ ID NO:12 orf2para

5 TCAGATGGCCAGGGCGGCGAAACCGCCGGACTGGAAGTCGTAGGCCACCGGTACCTCGATCAGGAACGGGCGGCCGAGTC  
CGGCGCCCTTGGTGAGGGCGGCGAGCAGCGAGGTGCGGTGCGTGCGGACGGCCTCGCAGCCGTTGGCCTCGGCGAGC  
TGGACGAAGTCGACGCTTCCGAAGCCGACGGCGGGGCGTGGGAGCGCTGGTGTCCGAGGTTCTGGTACAGCTCGATCAG  
GCCGTTGCGGTGCTTGTGACGACGACCATGACGATCGGCAGGCCACGGCGACGGCCGTCTCGATGTCCGGCGCTGTTGG  
AGTGAAGCCGCCGTGCGCCGCGATGAGGAAGACGGGCTCGCCGGGCGGGCGATCTGGGCGGCCATGGCGGCGGGCAGT  
10 CCGTAGCCGAAGCTGGAGCAGCCCGCGGAGGTGAGGAATCCGTACGGCTGGTGGACTTGGCGAAGAGCACGCCGTAGTG  
GCGGAAGAAGCCGATGTGCTGACGAAGGTGCCGTTGTGAGGACGGAGTTCATGCAGTCGATCACCTGGTGGACCCGCA  
TGCCGTCCTCGTACTCGGTGGGGTTCGGCGAGGAATTTCGGCGACGCGGGCGCGCAGGGCGCTGAGGTGCTGCCGGGTCTTG  
GGGCGAGGCCCGAGGTGCGCTGCTGAGCGCGGTGACGAATTTCGGCGACGTTGGTGACGATGTGATGTGCGGCGGGAA  
15 CAGCTCCGGGATCGGGTTGACCTCGGGGGCGACCCGACCGTGGTCTTGGCCCGGCCCGCGCTCCACATGGAGGGGCGCA  
GGTCTCGGGCTAGTCGTAGCCGATCGCCAGGAGGAGTTCGGCGGGGCGAAGATCTCGTCGAGGGCGGGTGGCCGAGA  
ATGCCGTCATGTAGCCGCTGATGGCGCCGTAGTTGAGCGGGTGGTCTGCGGCAGGACGCCCTTGGCGGTGTAGGTGGT  
GACGACGGGGATGTTTCAGCCGCTCGGCGAGGGCGCGCAGGGCGTTCGACGCCCCGCGCGGATGACGGCGCTACCGACGA  
CGAGGAGGGGTTCTCGGCCTCGCGCACCACTCAGCGGCTCGTCGAGGCGGGCGCGCCAGTTCGGCGTCCAGGGCGTGG  
GTGGCGGTGGCCCGGACAGGGGGGCGTGGTGGGGTGCCGTTAGTTCAGTTCGGCGCGGAGGAGTTCGCGCGGCGAGGTGAT  
20 GAAGCTGGGACCCAGCGGCTCGATCCGGCTGTTGAGGACGGCGCTGTGACGAGGTTGACGATGCTCTCGCCGCTTCGA  
GCTGGACGCTGAACCTGGTTCAGCGGGCCCATCACGGCGGTGCTGTCCAGGCACTGGTGGGTGACGTTGGGGTAGCAGTCG  
TAGACTCGGACTGCGCGGCCAGCGCGATGACCGAGTTCGGTCCAGGGCGGAGGTGGCGACGCCGTTGGCCAGGTTGGT  
CATGCCGGGGCCAGGGTTCGCGAAGCACGCTGGGGGCGGTTGGTGATCCGGGCGAGGACGTCGCCCATCACCCGCGCGG  
25 TGAACCTCGTGCCGGGTGAGGACGAAGTCAGTCTTCGACCTCGTCGAAGAGAATGGCGGACGCTCCCGGCCGACGACG  
CCGAATACATGGTCGACACCGTACTGGTGAAGACGTTCCAGCATGGCTTTCGCGGTGCTGGTGGCCAT

## SEQ ID NO:13 orf3para

TCATACGACCACCCGGCCCTGGAGCCTGAGCCTGCGCACCGCGTCGACGGAGCGCCGACCGTCTCGCCGAAGTCCACGT  
CCTCCGGCGGCACCGTGTGATGACCACCGCGTCTGTACAGGCGCCGTGCCATGGCGCCCTTGACGGCCGTACCTCGTCCG  
30 CGCCGGATCCCTTCGGCGAGGAGCAGTCCGGTCCACGCGCTGGTGGTGCCGGACCCCTCGTGGATGCCAGCTTGGGGCG  
GGCCACGGTCTCGGCGGGCAGCAGGCCGAGAGGGCTGCCGCAACACCCACTTGTGGTGCCCGCGGCGTTTGAGCC  
CGGGTTCGAGGGAGACCAGCGCGTCCAGGACCGCGGTCGTCAGTACGGGTGGTGGTCCACTTCCCGCGCATGCCCGCG  
AGGACGGGGACATCTCGTTGAGGCCGTGGAAGCCCGCCATGTGCCCCGCGATCTCGTCGTGAGGGACAGAGCGAGGC  
CGTGCGCCGGTGCATACCGCCGAGCGGGATGTGCGCGCCGTACCCGGTGGATGCGGAGCGGCCCGGTGTGAGCCGCC  
35 GGTAGAGGGCGACGAGCGGCAGCAGGTACTCCAGGACCGTGGGGTGGTGTCTCCGCGCGGCGACCGCCAGGGCAGT  
TCCCTGACGAGTTCGGCCGAGTGGAGCCGATCTCGTGTGCGCGGTGCCAGGTGGACGGCGACCGAGCGGGCGCGCTC  
GAACTCGTCGGACACCTCGGTGCCATCGACACGGACCGTGTCCCGGGTGCCAGGGCCCGCTGTGGGCGGCGACTCCCC  
CGGAGTCGATGCCGCCGACAGGACGACGGTGGGGGCGCCCTCCCCGCGCGCAGCCGGGTGCGGACCGCCGTGGCGAGG  
CGTTGCGCGACAGGTCCACCGCTCCGTTGCGCGGGCAGCGCCCGGAGAGCGGGGTGTCCAGGTGCGGACCGCCCT  
40 GCGGTTGATGTGCGGACCGCCGACTCCGTGACGAGGAGGCGGTCCCGCGGGGACCGGCGAGACCGCCCGCCCCCG  
GCGCGGTGTGGGTGCCGGACAGGCCAGCGGCCCGGCGCCGCTCGTGCGCCAGGGTCTTCGCTCGGTGGCGGCGCTCAGC  
CCCGTCAGTTCGGCGCGCAGCCACAGCGGTACCGAACCGGCTGGTGGTGGCGCGACGCTGCGGCGGAGGCGGTC  
GGTGAGCAGTGCGGCGAACCCTCCGTTTCAGGAGCGCGAAGGCCCGGGGCCAGCGCGCCAGGCGGCCAGCAGAGTT  
45 CGGCGTTCGCGAGGGCGGCGAGGAGCGCCGAGCGCTCCGTCAGCTCGGCGCGGTTGTACAGCTCGCCCGCCAGGAGC  
AGCCGGACCTGGCCGTGCGCGACAGGACGGCGGACGGCCAGGGTACGCGCCGTTCCGCTCCAGAGCGGGTACGCGGT  
GCCGTCGTGACGGGGACATGGGTCCCGCGACGGCGAAGCGGGGTGCGCTGCCGGGTTCGGAGTGACCGCGGGGGCCG  
CGCCGGGGCGGCCCTCGGTGCCGATGCGCACCCGAATCCGTACACGAGGTGGGGCCGGGCAT

## SEQ ID NO:14 orf4para

CTACCCCCACCGCTGCCCGGCGAAGTCCACGGCGCTCTCGGCGTCCACCGGTCCACCGGTTCTCGGCGTTCTCGGCGT  
CGTCCGCGCGCCCGCCCGGTGGCAGGGGAGAGTCCACCGGTGCCGACGCGGGCGACGTGGTGCGCGGGCGTACTGGTAG  
AGCAGTTCGGCCCCGATCTCCGCCGCCAGCAGGGAGGTGATCCCGACGGGTGTCAGCCGGGGACACCTCGACCACGTC  
GAAGCCGACGGGCTGAGCTGCCCGACACGTGACGAGGGTCAGCACCTCGCGGAGGACAGCCCGGGGGCGCGGTG  
50 TGCCGCTGCCCGGGCGTACGCCGGGTGACGACGTGTCGACGAGAGCTACAGCGGCGAGGCGCGGACGGTGC  
CGGATCTGCTCGGCGATGCCGCGCGGTGAGCGCGGGTGAAGTCGCGCGCGGTGACGATGCTGACGCCGTGCCCGCGCGC  
GTAGTCCAGGGAGTGGGCGCGGATTGTGGCCGCGGATGCCGACCTGGACCAGGCGCTCCGGGTCCACAGGCCCTCTT  
CGATGGCCCGAGCGAAGGGGTTGCCGTTGGTGGTAGGTGCCGCCGTAGACGGGTGGGTGGTGTGCTGTGCGCGTCCAGG  
TGCAGGACGGCGACCCGGCCGTGGCGGGCGTGCACGGCGCGAGGGCGGCCAGGGAGAGCGAGTGGTCCCCGCCAGCAT  
CAGGAACGCGTGGTTCGTTCCAGGACCGGGTCAGGGCGACCGTTCGCGGTGTCATCGCCGCGGACGAGGGGCG  
60 TGAGGTGATGTGCCCCCGTTCGAGACAGTCGATCCGGTTCGAAGACCCCTGGGCCCGGTGATGCCGACGCGGTGGATC  
AGGCTGGACTCGTGCCGATGGCGCGCGGCGCAACCGCGCGCGGGCGGGTAGCTGGTGCCTCCGTCGTACGGGGCGCC  
GACGACACACGTCATGGCCGATCGGGTGGGCGCGGTGGCGCAGCCGATGAAGGTGCGCGGTGGGCGTAGCGCGGGG  
AGACGGCGGTGGACAC

## 65 SEQ ID NO:15 orf6para

ATGCGTGCCCTTTCGCCAGAGGGTTCGCGGTGCACACGGTCACGCCGGATCAGGGGGTCCACGCGGACCTCGCCGT  
CATCGCCTCCGACGTTCCCGCGGCGGTGGCGCGGTGTTACCCGTTGCGGGTTCGCCGCGCGGAGTGTGCTGCTCAGCC

5 GGGACGCGGTGCGCCGACGGGATCGCCCGGGGCGTGGTGGTGTCTCGGCAACGCCAACGCCGGGACGGGCGCGGGG  
TACGAGGACGCCGCGGAGGTGCGCCATCTGGTGGCCGGGATCGTCTGACTGCGACGAGAGGGATGTGCTGATCGCCTCCAC  
GGGACCCGTGCGCGAGCGGTATCCGATGTCCCGTGTCCGGGCCATCTGCGGGCGGTGCGCGGGCCCTTACGGGTGCGCG  
ACTTCGACGCGCGCGGCGGCGCCGTGCTGGGCACCGCGGGCGCCGTCCACGATCCGGCGGGCGGGTGCAGCGACGCG  
10 ACGCTGATCGGTGTGCGCAAGGGCCCGGGTACGGGCGCGGCGGAGCAGGACGACCGGTGACGCTGGCGTTCTTCTGCAC  
GGACGCCAGGTGAGCCCCGTGTCCTCGACGACATCTCCGCCGGGTGCGGACCGCGCCTTCCACGGGCTGGGCTTCG  
GCGCCGACGCTCCACCGGCGACACGCGGCGCTTCTCGCAACGGGTCGCGGGCGGGTGGACCTCGTGCCTGTCGAA  
CAGGTCTGGGCGCGCTGGCGCTGGACCTGGTACGGGACGTCTGTCGGGACAGCGGTGCGGCGGCGCCCTGGTACGGT  
GCGGGTACCGGGGCCACGACACCGAGCAGGCGGGCGCGTGGGCGGGCGGTGGTGCAGCGCGCTGCTGAGGGCCG  
15 CGGTGCACGGCCCGCACCAGCTGGGCGCGCGTGGCGCGGTGGCGGGTGGACACGGGGACGAAGCCCCGGCCGGTCT  
CCCGGGCGGATCAGATCCGGGTGCGCGGCGGGAGGTCTTCCCCGCCCCCGCGACCGGGCCCGCCGACGCCGTAC  
CGCGTATCCGACGGCGGCGAGGTGACCGTCCATATCGACCTCGGTGTCCCGGGCGGGCGCGCGGCTTACGGTCC  
ACGGCTGCGACCTCCTGGCGGGGTACCGCGCCTCGGCGCCGGCGGGCGCTTGA

## 15 SEQ ID NO:16 para cluster

CCATGGGAGCAGCATCGCAGTGCCTCCCCGGCGCCATGCCGCTAGCTGGTAGTCCCCCTGCCGGGTGCCGACCGCCG  
GGGCGGTCCCAGGTGCGGCGGCCGATCTAGTGGTGTCTCCGACGGTGCTGCTGGGTGAGGGGAGTGTGAGGCGGA  
TGGTGGTTCCCGCGCGGGCGGGCTGTGACGCGCAGTTGGCGCCGAGTGCTCCACCCGGTCGGTGAGGCGGACGAGG  
20 CCCGAGCCCCGCGAGGGGCGGCGCCACCGCGGCGCTGTCGCGATGCCGACGTGGAGCCGTCCGTCGGGTGGCCAC  
ATGGACGTGACGACGGTGGACCGGAGTGCTTGGCGCGTGGTACGGGCTCGGAGACGGCGTAGTACGCGCGGTCT  
CGACCGGTTCCGGGTGGCGTTCCCCGGTCTGGATGTGAGCCGACCGGATGGCGGAGCGCGGGCCAGGGCCTTGAGC  
GCCGGGCGGAGTCCGCCCTCGGCGAGTACCGCGGGTGGATGCCCGGGCGACCTCCCGGAGTTCGTGACGCGCGCGC  
CAGCCCGTGGTCACTCGTTCGAGTCCCGGATCAGTCTGTCGGCGTGCAGCGGCACCGACAGTTGCACGGTGGCGACCC  
25 GCAGCGCCAGGAGACGAGCGCTGTTGGGGCGCTGTCGAGTTCGCTTCGATACGGCGGGCGGTGTGCGCGCG  
GCGACGATCCGGGCGCTGACGCGGTGACGCGCGCTGCTTCCCGCTGGCGATGGCGGTGGCCACCGAGTTCGGTGAA  
GCCGGCCAGCGGTCTCGGTGTCCGACGGCATCGGCTTGTCTCATCGACGCCAGCTGAGCGCGCCCCACAGTTGTC  
CGTCGACGTTGATCGGCATGCACACCGTGGCGCGGAATCCCCACTCCTTGCCGACGACGAGGCGCGGCCGAGGACAG  
GCCGCGTAGTCTGTCATCCGCGCGGGCAGCCCGACTCGAACACAGGGTGTGCACATTCCGGCGCGCGGGCGGTACCTG  
GATACCGGCGGGAATACGCGCGGTCTGGTCCAGGCGGCGACATACAGGGCGGTTCGGTGGGCTCGTAACGGCCGA  
30 GGACCGCAAGTCCGGCGAGAGGAGCTGTCGGGCTCGCGCGGACCGCGCGCAACACCTCCTTCGCGGTGCGCGCCG  
GCGACAGGGTCCGACGCGCGCAGCGCGCTGCTCCTCGGCGGCCCCCGCAGCTCCACACGTGCTGGGTGTTTCG  
GATGGCGGTGGCCACGAGGTGGTGAACCGGCCAGCCGTCCTCGGTGTGCGGCGGCGAGCGGTTCGCGGTGAGCGAGA  
TCGCCATCATACGCCCCACAGCCGTCCCTCGACGTTGATCGGCACGCCGACGACCGAAGCCGCGCGCCCTGGCG  
AAGTCCGCGGGTCCCCCGGACGACTCGGCGGCGTCTGATCCGGCGCGGCGCCCCGTCTCGGACACGAGCTCACCAC  
35 GTTCCGCGCGTCCGGGTCCACCGGGTCCAGCGGAGAGCGGGCGGTGCGAGACTTCTGGACACGCGCGGCGCGC  
TCGCCATGCCGTCCGGATCGAGCCTGATGATTCCGGTCACATCGTTGCCGAGCAGTTCTCCGACTTCGCGCGGCGACCGTC  
GCGAACATCTGTTCCGGTGGGGTGGCCCTGGCCACAGGGTCCGACCCCGTCCGAGTGCCGCGCGCTCTCGACGATCTG  
TTCGACGACACGACCGCTGCCAGGCCCCCTACCCGCGCGATGACGCGCGCATACCGGGTATACGCGACATCAGCATG  
ACGTCCGCGGTGAACGCCGTCAACGTGGCCCGCGGAGTCGGGAACGCGTCCGGAATCAGCCCGGAACGGCGGGA  
40 CCGTCTTCTCCGTCCGCGCGGGGCACTGCGCGCGCGGAATCCGCCCTGACCTCGGGAGTTTCGAGCTAGCTGGAAT  
CAGCGTTTCGGGTGGTGGGAAGGATGTTGGCGCTGGCGCGATGCGGAAGCCGATCGTTCCAGTACTTCTGGGAAG  
TGCGTTCGCGGAGAGTCCGTCCGCTTCCCCGAGTGGGCGCGACGACGCTGCGGGTCTCCACGGGGGAGAGATCCGCGAA  
CCGGCGAAGGAGTGCCTGTGCGACGTCTTCGATCCGAGAAGAGTTCCGCCGTGTCCGGACCCGCGCGGCAACGTCC  
CCACCGCGCTCTGTATCAGCGCGCTCGGCGCGCTCAGCCACGAGAGAAGATCGGATACGAGTGTACGAGTGCAGCGA  
45 TGAGGTTCTGTCAGCAGTCCCCGGCTCCGGGTCCGTACCGTCCATCACCCTCCTGGGCTGTCTGGGCGTACGCGCCG  
ACGGCGGAACTGGAGCTGGGCCCTCGCGTACGCGGCGGTTTTCGCCCTGCTGCTCATCAACCGGGCGAGTGTGGTG  
CCGGTTCGACTCGATCGTCTTCCGTATCTGGGCAACTCACCACGGGCGCGGTACCGCGACGCTCCAGTCTATGTGTC  
CCGGTTCGCGAACTCCTGGCCGAGTGTGTCTCCCGACGTTTCGACACCCGAATGCTGCACAGCCGCGGGCTACA  
CCCTCGCGCTCGGCACCGAGCAGATCGACGCGAACCCTTTTTCGAGAGGCGATCAGGACAGGGCGCGGCTCTCGCGGAG  
50 GAGCAGCACCAGGAGGCGGGCGCGTGTCTGCCAGGCCCTGCTGAGCTGGGGCGGGACACCGTACGAGGAGCTGAGCGC  
GTACGACTTCGCGCTCCAGGAGGCAATCGGCTGGAGCAGCTCCGGCTGGGCGCGGTGGAGACATGGGCGCACTGCTGT  
TGCGGCTGGGGCGGGACGAGGAGGTGATGGACAGCTCAAGCCGAGGTGCAGCGCAATCCGCTGCGGGAGCGGCTGATC  
GGGCGCTCATGACGGCGAGTACCGGTGGGGTGCCAGGCGGACGCGCTCAGGACGTACGAGGCGACGCGCGGGCCCT  
GGCCGAGGAGTGGGGACCGATCCGGGCAAGGAGTGGCGGCGGTGACAGCGCGGATCCTGCGTACGACACCGGTCTGG  
55 ACCGCGTCTCCCGCGTCCGCGCGCGCTCGGCGGGGTCGGGCGGGGCGGTGACGGTGTGCGTCCCGGCGACGCGG  
TCAGGCGGTTGACGCGCGGCGGTGGCGGGGCGCGGTCCGCGGGCGATGACGTTGGGCGGGCGGGGCGGGCGCG  
CCCCGCTCCGCTCCGGTCCGTTTCCGCGTCCGTTTCCGGCTCCGGCTCCGGCTCCGGCTCCGCTCCTGCGTCCGTTT  
CCACCTTCTTCCCGCTCCGTTTCTGGCTCGGCGTCCGTTGCCGCGTCCGTAGCCGCGCCGTTTCCGGCCATGTCTCC  
GGGCGCGGTCCGCTTTCGGGTCCGTGGCGCTCCACCGGCGCGAGACCTCCGGGGCGAGCCGGTCCACGGGGCGCGCA  
60 GGGGATGCGCACCGGGCAGGTGTTCCACGCTGCCGCGGTTCTGTCGGGCGCGGCGACGAGCTGCGCGGTCTGCTGGAGT  
CCGCGAGTCCGCGTTCCACACTCGGGCGGGTGGCGCTCTGTCGCGGAGGCGGGCAGCGGCAAGACCCGCTCCTC  
TCCGAGTTGGAGCGCTCCGTTCCGACAGTGTGCGCACCGTCTGCGCGTCTGTTTCGAGAGTGAGGACCGGCGGACTA  
CTGGCGGTGGACGACCGTGTGCGGCATCTGTACGCGATGTGGCGGAACGTATGCACGGATTCCCCGTTGGTTCGCGC  
GCGCACTCGCGGAACGTCTCCGAGGTGGGCGCGGAGCCACAGGGGCGCACTCCCCGACGGGGCGAGGAGAACAGC  
65 GGCAACGGGGACGGTGGGGCGACGGGGACAGCACCCCGCGCACACCTCAGCTCGCGCCCGCTCTCGCGCCCCCGCG  
CTCCAGAGAGGCTCGTTTACCCTGCACGACCGCTGTGCCAGGCGCTTCTGCGACGGTCCGCGAACCCGTGGTGTATCA  
TGCTGGAGGACATGGAGCGGGCGGCGGCTCGCTCCGCTCCTGGTGGAGCACTGCGCACCGTCCCG  
CTGCTGCTCGTGGTACACGCGCACCTTCCGGTTCGCGCACGACCGGAGCTGCGACGGGCGCGCGGTGATCTCCA

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

GTGACCGGGCGCGCGCCGGTCTGCTGAACGCCCTGGACGCACGGGCCACCGGGGAACCTCGCCGGAGGGATGCTGGGCA  
AGGCCCCGGACACCCTCTCTGCTACGGGCCCTGCACGAGCGCTCCGCGGGGAACCCGTAATCTCTCTGCTCCAGCTCTCCGC  
TCGCTCCGGCAGGGGCTCGCCGCCCTGGGAGACGGAGATCCCGACGAGCTGGCCGGGGTCTGCTGCAACGGCTGTC  
GAGCGTGCCGCCCGCGTGCGCCGGGTGCTCGACATCTGCGCGGTCTGTGGAGCGCAGTTGCAACGGCGTGTGATCGAGA  
CCGTGCTGCGCCATGAGGGAATCCCGCTGGAGAACGTCCGTACGGCGGTCCGCGGGCGGTCTGCTGGAGGAAGACCCCGAC  
GACCCCGGGCGGCTGAGGTTCTGTCATCCGCTGGTCCGGGAGGCCGTCTGGGACGACCTGGAGAACACCCGTCCGCCCGT  
GTCCCGTTCTCCGCGCTCGGGGCGCTGGCCACGGTCTGAGTCCCGGGGCCCGGGTCTCTCGCGCGGCGGCGCGCTTGC  
GCGCTCCCGACGCCGGGCTTGATCCCGCGGGCAGCCGACGCGCAGCCGGGTGCAAGGGGCGGTGCCGACACTGGGCG  
GGCGGCGGCCGTGGCCGGTCCGCCCCCCACGGCCACCAGGAGCCCCATTTGGACACGTACGACGCGGATACGTACC  
CGCGGTCCGGCACCCACCGAGCCGCTCCCGACGCACCTCCCCACGCGCTCCCGGGACCCGTCCCGGCACCCGTTC  
GAGCGCGCCCGGACCCGGGCGCCGAGGCCGCTGGTCTGCTCGCGCGGACCCGCGCCATATGTTCCACCCGGTCTCTGCC  
CCGGGCGCGGAGGACCGCTTCTGGTCTCCGGCGCGGTGACCGGTACGGGACACCGAAGGGCGCACCTATCTCG  
ACGCTCTGTCGGTCTCGGACTGACCCAGATCGGCCATGACGTGAGGAGATCGCGCAGGCCCGCCGAGCAGATGCGG  
ACACTCGGTCACTTCCACACCTGGGGCACCATCAGCAACGACAAGGCCATCCGACTGGCGCGCGCTCACCGACCTGGC  
GCCCCAGGGTCTCCAGCGCTTACTTACCAGCGGCGCGGCGAGGGCGTGGAGATCGCCCTGCGCATGGCCCGTTACT  
TCCACCACGCGACCCGCGAGCCCGGAGCGCACCTGGATCTTGTGCGCGCGCACCGCTTACCACGGCATCGGCTACGGCAGC  
GGTACGGTGTGGGCTCGCCCGCTTACCAGGACGGGTTCCGGCCGGTGTGCCCATGTGACACACCTCACGCCGCCCGA  
CCCGTACCACGCCGAGCTGTACGACGGCGAGGACGTACGAGTACTGCCCTGCGCGAATCGCCCGCACCATCGACGAGA  
TCGGCCCCGGGCGGATCGCCGCGATGATCGGGGAGCCGGTTCATGGGCGCGGGCGCGCGCTCTCCCGCCCGGACTAC  
TGGCCGCGCGTCCCGCGCTGCTGCGCTCCACGGCATCTGCTGATCTGGACGAGGTCTGTCACCGGTTCCGGCCGAC  
GGGGACCTGGTTCGCGGCCGAGCACTTCGGGGGTGACCCCGCATCTGCTGGTACCGCGAAGGGCATCACTCCGGGTATG  
TCCCGCAGGGGGCGGTGCTCTGACCGAGGAGTCCGGGACGCGGTGAACGGGGAGACGGGGTTCCCGATCGGCTTACC  
TATACCGTCAACCCACGGCGTGCCTGCGCTGCGCAATCTGCACATCATCGAACGGGAAGGGTGTGAGGAACGC  
GGTGAAGGTGGGCGACCACTCGCCGGGCGGCTGGCGGGCTGCGCGGGTGCCTGGCGGCGGCTCCGGCAACTGG  
GCATGATGCTCGCCGTGAGCTGGTGTGCGACAAGACGGCCGACCCCGTGCCTGGGCGGCGACCCCTCGGGGTCTGGAC  
GCGCTGCGCGAGGACGCGGGCGTTCATCGTCCGGGCCACGCGCGCTCCCTGGTCTCAATCCGGCGCTCTGTGATGGACCG  
GGCCACGGCGGACGAGGTGGCGGACGGGCTGACTCGGTGCTGCGCGGGTGGCACCCGACGGGCGGATCGGCGCGGGCCC  
CCCGCGGGGGTACGAGACCGGGCCGCCACCCGCGGGGGCGCGGGTGGCACAGCGGCGGACCCGGCGCTTCCC  
CGTTTCCCGCGCTTTTCCGTGCCCGCGCGCTTCCCTGCGCCCTGCCCCCTGCCCCGTGCGGCGCTCTCTCTCC  
GCTGTGGCGCGCTTCCCGTTCAGCGCGCTGTGAGCCGCGGCCAAGCGCCCCGTGCCACGGTGGGAGACCGCGCCCGA  
CGGGGCGCGCGGAGCCGGCAAGCCGAAGGGAAGTCCCGTCCGATGCGTGCCTCTTCCGCCAGAGGGTTCCGCGTGCACC  
ACGGTTCAGCCGGGATCAGGGGGTCCACGCGGACCTCGCCGTTCATCGCTCCGACGTTCCCGCGCGGTTCGGCGCGGTG  
TTCACCCGTTCCGGTTCGCGCGCGCGAGTGTGCTGCTGAGCCGGGACGCGGTTCGCCGACGGGATCGCCCGGGGCGTGGT  
GGTGTCTCCGGCAACGCCAAGCGCGGACGGGCCCGCGGGTACGAGGACGCGCGGAGGCTGCTGTTGGTGGCGG  
GGATCGTCTGACTGCGACGAGAGGGATGTGCTGATCGCTCCACGGGACCCGTTCGGCGAGCGGTATCCGATGTCCCGTGT  
CGGGCCCATCTGCGGGCGGTGCGCGGGCCCTTACCGGGTGCCGACTTCGACGCGCGGGCGGGCGGCGGTGCTGGGACCGC  
GGGCGCCCGTCCACGATCCGGCGGGCGCGGTGCGGCGACGCGACGCTGATCGGTGTGCGCAAGGGGCCGGGTACGGGCC  
CGGCGGAGCAGGACGACCGGTGACGCTGGCGTCTTCTGCGACGAGCCGAGGTGAGCCCCGTCTCTCGACGACATC  
TTCGCGGGGTTCGCGGACCGCGCTTCCACGGGTGGGTTTCGGCGCGACGCTTCCACCGCGACACGGCGCGCTTCT  
CGCCAACGGGCTCGCGGGCGGGTGGACTCGTTCGCTGCGGTTCGAACAGGTCTTGGGCGCGCTGGGCTGCTGGTACGG  
ACGTCTGTCGGGACAGCGGTGCGGCGGCGCCCTGGTACGGTTCGGGTACCGGGGCCACGACACCGAGCAGGGCCGGG  
CGCGTGGGCGGGCGGTGGTTCGACGCGCGCTCGCTGAGGGCGCGGTGACGCGCCGGCACCCGACTGGGCGCGGTGCG  
CGCGTGGCGGGTGGACACGGGACGAAGGCCCGCGCGGTCTCCCGGGCGGATCACGATCCGGGTTCGGCGGCGGGGAGG  
TCTTCCCCGTCCTCCGACCGGGCCCGCCGACGCGCTCACCGGTATCCGACGCGGCGAGGTGACCGTCCATATC  
GACCTCGGTGTCGCGGCGGGCGCGCGCTTCCAGGTCACGCTGCGACGCTCCTGGCGGGTACCCGCGCTCCG  
CGCCGGCGGGCGGTCTGAACGGGCGCTCCCGGGCGGACGGCGACCGCGAGGGCGCGGGAGCGCAGGGAACACGGGAGCG  
GGCCCGGTGGTTCGATCGGCCACCGGGCCCGTCCCGTCTGCTCCGTCGCTGCTCCCGGGCGGCTTACCCACCGCTGCC  
CGGCGAAGTCCACGGCGCTCTCGGCGTCCACCGCTCCACCGGTCTCTCGGCGTCTCGGCGTCTCGGCGCGCGCCCC  
GGTGGCAGGGGAGGTCCACCGGTGCGGACGCGGGCGACGTGGTGGCGCGGGCGTACTGGTAGAGCAGTTCCGGCCCCGAT  
CTCCGCGCGCAGCAGGAGGTGATCCCGACGGGTCTGACGCGGGGACACCTCGACACGTGAGCCGACGGGCTGA  
GCTGCCCCGACACGTGAGCAGGGTACGACCTCGCGGAGGACGCGCGGGGGCGGGTGTGCGGTGCCCGGGGCG  
TACGCGGGGTGACGACGTGATGTGACGAGACGTACAGCGGACGGCCGCGACGGTGCCTGGATCTGCTCGGCGAT  
GCCGCGCGGTGAGCGCGGGTGAAGTCCGCGGCGGTGACGATGCTGACGCGGTGCCGCGCGCTAGTCCAGGGAGTCCG  
GCCGCGGATTGTGGCCGCGGATGCCGACCTGGACAGGCGCTCCGGGTCCACAGGCCCTCTTCGATGGCCAGCGGAAG  
GGGTGCGGTGGTGGTAGGTGCCGCGTAGACGGGTGGGTGGTGGTGTGCGCGTCCAGGTGACGAGCGGCGACCG  
GCCGTGGCGGGCGTGCACGGCGCGCAGGGCGCGGAGGAGAGGTGGTCCCGCGCCAGCATCAGGAACCGTCTGTTG  
GTTCCAGGAGCCGGGTGAGGGCGACCGTTCGCGGTGTCCATCGCCAGGTCCATCGAGAAGGGGCTGAGGTGATGTGCCCC  
CCGTGCGACACGTGATCCGGTGAAGACCCCTGGGCCCCGCTGATGCGGACGCGGTGGATCAGGCTGGACTCGTGCCG  
GATGGCGCGCGGCGGAACCGCGCGCGGGCCGGTAGCTGGTGCCTCCGTCTGACGGGGCGCGACGACACACGTTCAT  
GGCCGATCGGTCGGGCGGTGGCGCAGCGCATGAAGTTCGCGGTGGGCGTAGCGCGGGGAGACGGCGGTGGACAC  
CTGGCCGTTCCCGCGCACCCCGGCTGCTCCGTTCCGTTCCGTTACGACGCGCGGCCACCCCGTCCCGGTTCCCGT  
GCCGACCCCGTTCCCGAACGGGCTCCCGTTCCCGGTGGAATCCCGTTCCCGCGCGCGCGGGCGCGGTCCGGGCGCGG  
TGCCCCCTCCCTCCGAGACGCTCTGCGGTTCTGCGGCGGTGCGGCTCTGCGGGCGGGTGGCCGCGCCACGCCCCGCT  
GCACCGTCCGCGCGCGCGCGGTGCGGTTGCGCGCGCGGTGCGGTTCTGGCCACCGGTGCGGTTCTGCGCGCTCATACG  
ACCACCGGCCCCGGAGCCTGAGCCTGCGCACCGCGCTGACGGAGCGCGCACCGTCTCGCCGAAGTCCACGTCTCTCGG  
CGGACCGGTGTCGATGACACCGCGTGTACAGGCGCGCTGCCATGGCGCCCTTGACGGCCGTACCTCTGCGCGCGGA  
TCCCTTTCGGCGAGGAGCAGTCCGGTCCACGCGTGGTGGTGGCGGACCCCTCTGGATGCCAGCTTGGGCGGGGCCAG  
GTCTCGGCGGGCAGCAGGCGGAGAGGGCTGCCGAACACCCACTTGTGCTGCCCCGCGGCGGTTTGGACCGGGGTTT

GAGGGAGACCAGCGCTCCAGGACCGCGCGGTCCAGTACGGGTGGGTGGTCCACTTCCCGCGCATGCCCGCGAGGACGG  
 GGGACATCTCGTTGAGGCCGTGGAAGCCCGCATGTGCCCCGCGATCTCGTCTGAGGGACAGAGCGAGGCCGTGCGC  
 CGGTGCATACCGCCGAGCGGGATGTGCGCGCCGTACCCGGTGAGGATGCGGAGCGGCCCGGTGTCGAGCCGCCGGTAGAG  
 5 GCGCAGGAGCGCGCAGGTAATCCAGGACCGTGGGGTGGTGATCTCCGCGCGCGGACCGCCAGGGCAGTTCCTTGA  
 CGAGTTTCGGCCGAGTGGAGCCGGATCTCGCTGTGCGCGGTGCCAGGTGGACGGCGACCGAGCGGGCCCGCTCGAACTCG  
 TCGGACACCTCGGTGCCATCGACACGGACCGTGTCCCGGGTGCCAGGGCCCGCTGTGGGCGGCGACTCCCCCGGAGTC  
 GATGCCCGCGGACAGGACGACGGTGGGGGCGCCTCCCCGCGCGCAGCCGGGTGCGGACCGCGTGGCGAGGCGTTGCG  
 CGACCAGGTCCACCGCCTCCCGTTCCCGGGGAGCGCCCGGGAGAGCGGGGGTGTCCAGGTGCGGACCGCCCTGGCGGTG  
 10 ATGTGCGAGCCGCCACTCCGTGCAGCAGGAGGGCGGTCCCGGCGGGGACCCGGCAGACGCCCGCGCCCGCGCGCGGT  
 GTGGGTGCCGACAGGCCCAGCGCGCGGCCCGGCTCGTGCGCCAGGGTCTTCCCTCGGTGGCGCGCTCAGCCCCGTCA  
 CGTCCGGCGCGCAGCCACAGCGGTACCGAACCGCGGTGGTGGTGGCCGCGACGGTCCGCGCCGTGGAGGCGTGGTGAGC  
 AGTGGCGCGAACCGTCCGTTCAGGAGCCGGAAGGCCCGGGGCCAGCGCCGCCAGGCGGCCAGCAGAGTTCCGGCGTC  
 GCGAGGGCGCGCAGAGGAGCCGAGCGCTCCGTGACGTCCGGTGGCGCGGTGTACAGCTCGCCGCCAGGAGCAGCCGGA  
 CCTGGCGCTCGCGCAGGACGAGCGGGCGGACCGGCCAGGCTCACGGCCGTTCGCTCCAGAGCGGGTACCGGTGCCGTG  
 15 TGCACGGGGACATGGGTCCCGCGGACGGCGAAGCGGGTGCGCTGCCGGGTTCGGAGTGACCGCGGGGCGCGCCCGGG  
 GCGGCCCTCGGTGCCGATGCGCACCCGGAATCGTACACGAGGTCCGGGCGCGGCATGGTGAATCGTCTCCACGGTGG  
 TCAGATGGCCAGGGCGGCGAAACCGCCGACTGGAAGTCGTAGGCCACCGGTACCTCGATCAGGAACGGGCGGCGGAGTC  
 CGGCGCCCTTGGTGAGGGCGGCGAGCAGCGAGGTGCGGTGGTGGCGCGGACGGCCCTCGACCGCTTGGCCTCGGCGAGC  
 TGGACGAAGTCGACGCTTCCGAAGCCGACCGCGGGGGCGTGGGAGCGCTGGTGTCCGAGGTTTGGTACGCTCGATCAG  
 20 GCGGTGCGGTCTTGTGACGACGACCATGACGATCGGCAGGCCAGGCGCACGGCCGTCTCGATGTGCGCGCTGTTGG  
 AGTGGAAGCCCGCTCGCCCCGATGAGGAAGACGGGCTCGCCGGGCGGGCGATCTGGGCGGCCATGGCGGCGGGCAGT  
 CCGTAGCCGAAGCTGGAGCAGCCCGCGGAGGTGAGGAATCCGTACGGCTGGTCCGACTTGGCGAAGAGCACGCCGTAGTG  
 GCGGAAGAAGCCGATGTGCTGACGAAGGTGCCGTTGTCGAGGACGGAGTTTCATGAGTCGATCACCTGGTGGACCCGCA  
 TGCCGCTCTCGTACTCGGTGGGTGCGCGAGGAATTCGGCGCAGCGGGCGCGAGGGCGCTGAGTCTGTCGGGTCTTG  
 25 GGGGCGAGGCCCGAGGTGCGCTCGTGCAGCGCGGTGACGAATTCGGCGACGTTGGTGACGATGTGATGTGCGCGCGGAA  
 CAGCTCCGGGATCGGGTTGACCTCGGGGCGACCCGACCGTGGTCTTGGCCCGCCCCGCGTCCACATGGAGGGGCGCA  
 GGTCTCGGCGTAGTGTGAGCCGATCGCCAGGAGGAGGTCCGGCGGGCCGAAGATCTCGTGCAGGGCCGGGTGGCCGAGA  
 ATGCCGTCCATGTAGCCGCTGATGGCGCCGTAGTTGAGCGGGTGGTCTGCGGCGAGGACGCCCTTGGCGGTGTAGGTGGT  
 GACGACGGGGATGTTACCCGCTCGGCGAGGGCGCGCAGGGCGTCGACGGCCCGCGCGGATGACGGCGCTACCGACGA  
 30 CGAGGAGGGGTTCTCGGCCCTCGCGCACAGCTCAGCGGCCCTCGTGCAGGCGGGCGCGCCAGTCCGCGTCCAGGGCGTGG  
 GTGGCGGTGGCCCGGACAGGGGGCGTCCGTGGGGGTGCCGTTGAGTCCGCGCGGAGGAGGTGACCGGCAGGCTGAT  
 GAAGCTGGGACCCACGGGCTCGATCCGGCTGTTGAGGACGGCGCTGTGACGAGGTTGACGATGTCTCGCCGCGTTTCA  
 GCTGGACGCTGAATTTGGTCAAGCGGGCCATCACGGCGGTGCTGTCCAGGCACTGGTGGGTGACGTTGGGGTAGCAGTCG  
 TACGACTCGGACTGCGGGCGAGCGCATGACCGGCTCGGTCAGGGCGGAGGTGCGGACGCGCGTGGCGAGGTTGGT  
 35 CATGCCGGGGCCAGGCTCGCGAAGCAGCGCTGGGGGCGGTTGGTGTGATCCGGGCGAGGACGTCGCCATCACCCGCGCG  
 TGAATCTGTCGGGTGAGGACGAAGTCGAGTCTTTCGACCTCGTGAAGAGAATGGCGGACGCCCTCCCGGCCGACGACG  
 CCGAATACATGGTGCACACCGTACTGGTGAAGACGTTCCAGCATGGCTTTTCGCGGTGCTGGTGGCCATGGAGATCTCCTT  
 CGCATCGGACGGGCGCGGGGATGGCGCCCCGGAACCGCGGACCGGGCGGTGCGCACCGGGTGGCGCACACCGTGGGTG  
 GTGGCGTTGCCACTGTGCGGATCGCCTCTTGGCGGCGGTGCGACGCCCGGCTTGGACAGAATGGGCAAGGCGCGTTCAAG  
 40 GCATGGCGTCCATCGTCTCGTGGCGCTTTTCGTGAATCCGTCCGGCGCGGAGGTCCTCATCCGATTCCGTCCCCTTC  
 CGTCCACCGATCCGAGGAGAATCCATGGATGTCTGGCGCGGTTGGAGCGCAAGCCAGCCTGAATCTTTTCCCATCGA  
 GAACCGGCTGTGCGCGCGCGCCAGTGCCGCGTGGCCACCGACCGCTCAACCGTATCCGTACTCCGAGACCCCGGTGG  
 CCGTCTACGGCGATGTACGGGGCTGGCCGAGGTGTACGCGTACTGCGAGGACCTGGCCAAGCGCTTCTTCGGGGCGCGC  
 CACGCGGGTGTGACGTTCTGTCCGGTCTGCACACCATGCACACCGTGTGACCGCCCTGACCCCGCCGGGCGGGCGCGT  
 45 CCTGTCTCGCGCGGAGGACGGCGCCACTACGCCAGGTGACGATCTGCCGGGGTTCGGCTACGAGGTGAGTTCT  
 TACCTTCGACCGCGGACACCTGGAGATCGACT

## SEQ ID NO:17 cvm cluster

GGTACCGGCATCCGACCCAGGCCCGGGCGCAGGACCCGGAGGCAGGCACCGGCACACCC  
 50 CGGCCGGGCGGCCCGGCTCCCGGCGGTGGTGTCGGGCGACCCGCAATCGGCAGCCGCC  
 CAGGCCCCGGACAGGAGCCCGGCTCAAGGACCCGGCCCTGCGCACCCGCTGAGGCGGCAG  
 GTTCTTGACAGCCGCGCATCCGCCAGTCCGCGCGGGGCGAGCCGCCAGGCGCCCGGCCG  
 GCACACCCGTGCGAGCGCCCGCTCCCGGCGGTGGTGCCCGGAGGCGGCGACCGGCAG  
 55 CCGGACACGGCCCCGCTCGGGGCGCGGCCAGGCGACAGGCCCTGGGACCCCGCTCGGAC  
 GCCCCGTTCCGACAGCAGGCCCGTGGGAAGCCCGCGGTGAGGCCCGCAGGCAGCCACCGGT  
 CGGCGGGCGGATCAGGTGTTGGCGGGGACTCGTCCGGGAAGATCTTTGTGACGACGGTC  
 CCGTCTCGGTGAGATAGCCGTGCAGCATCCCGGGGCTGCTGTGCGGCGCGTGAAGTCG  
 CCGCGGGGTCGAGGGCGATCACGCCGCCCTGCCCGCGAGCCGGGCGAGGCGCTTGACG  
 60 ATCACCTCGTAAGCGGGGACGCCACCGCGAGCCCTTGAATCTGATCAGATGGGAGAGG  
 GTCGAGGTCCGCGCGCCCGGATGAACACCTACCGGCGCGGTTGGCGCTCGCGCGACG  
 GTCCGGTTGTGCGCGTAGGTCCCGGCCCGATCAGCGGGGAGTCGCCGATCCGGCGGGG  
 AGCTTGTGTTGAGCGCCCGCGGTGGAGGTGGCCGCGCGAGATCGCCCGCGCGGTGAGG  
 GCCACCGCGCCACCGTCCCGTGCAGTGCAGGTGGGCCAGTGCTCCGGGGCCCTCCGG  
 65 GCGGCGGATCGCCGCTCGTCTCTCGCGCGCAGCAGCGCGTCCAGCGGGCCTGG  
 GTCCAGTAGTAGTCTGGGTGACGCTGCGCAGCCCGTGGCGGGCGCGAAGTCGTGCGCG  
 CCTCGCGGAGAGGAGGACGTGCTTCGACTTCTCCAGCACCAGCCGGGCGCCCTCGACC  
 GGGTTGCGCAGGAGGTGACCCCGCGCAGCGCTCCCGCTTCAGATCGGAGCCCCGCATC



ACGGAGGCGTCCAGCTCATGCCCGCGTCCGCGGTGAAGACGGCGCCCTTGCCCGCGTTG  
 AACAGCGGGTTGTCTCCTCAGTTCCGCGACGGCGCGCTCGACCGCGTCCAGGCTGTCCCG  
 CCGCGCGGAGCACCCGCTGTCCGCGCGGAGCGCTGCGGCGAGCCCGTCCCGGTACGCC  
 TTCTCCCGTTCGGGGCCGGTCTCTCCCGTCCAGGGCGGCTCCGGCCCCGCGTGGACG  
 5 GCGATGACCACGTACGGGCGTCCGGCCGGGGCTTCCCCGGCGCGCTCCCCCGTTCCCTC  
 TTCTCCTCCCGCGCCTGTCTCTCTCTGCTGCGTCTGTGGGCGCGCGGTGGGT  
 CCATGGCGCGCCGAGGCCCGGGTACGACGATGAGCGTGGTCTGTCAGCACCGCGCGCGG  
 AGCAGGGAGGACGCCAGCCAGGCGGTGGCGGGGCGGTGGGGCATCGGGCACTCCTCGGGA  
 CGGGGGTGAGAGACGCTCCGGCCGACTGTACTGACATGCCCATGCCCCCTCTAGTGCCCC  
 10 GGAGCCGCTTCCGCGCTCCCGCGCGCCGCGCGCCCGCGCGCGCTCAGTCCAGG  
 GCCAGGTCTCCGGGGCGGAGCGGGCGAGTCCGGCGAGTGTGCCGAGCGCCCGGGTCACT  
 TCGTCCGCGGACGGCGAGCCAGGCCAGCCGACCGCGTGGGTGTACGGCCCTGCCCG  
 GCGCAGAAAGCGCGCGGGCGTCAACCCGATCCCGTCCCGCGCGCGCGCGCGACGAAG  
 GTGTCCGGCGCGCAGGGGGCGGGCAGCACCCAGCAGTGGTACGAGCCGGGGTCCGCC  
 15 GACACGGCGAAGCCGTGAGCGCGCGCCGGGCGATCTCTGCGGTACGCCCGCGTCCCGC  
 CGCTTGCGCGGTACAGCGCGTCCAGCGTCCCGTCTGCGAGCGGACCGCGCCCTCC  
 AGCGGAACCGCGCGGGGGCGAGACCGCGGAGCGCAGCGCGCGCGCGACCGCTCCGTCC  
 AGCCCCGGGGCACCACCGGAACCCAGGGTCCAGCCGGGGGCGAGCCGCTTGAGAGG  
 CTGTCCGACGAGCACCGTCCGCCCGGGGGCGACCGCGGAGCGGAGCGGTGCCCTCCCGC  
 20 AGGAAGCCCCAGACGGCGTCTCGACCGCGGGAAGGTCCAGCCGCTCCAGGACCGCGCG  
 AGCTGGGCGAGACGCCCGTCCGACAGGGTGAAGGAGAGCGGGTGTGACAGGGTGGGCTGG  
 ACATAGACCGCCCGGAGCGGAGCGCTCCGGTTGGCTCGTCCAGCGCTCCGGAATCACC  
 CCGTCCGCGTCCATGGCGAGGGGACGAGCGTGATGCCGAGCGGGCGCGATCGCCTTG  
 ACCACGGGGTAGGTCACTCTCGACCCCGAGTCCGCCCCCGCGCGCACAGCGCGCGG  
 25 AGCAGCGGAGAGTGCCTGCCGACCGTTGCCCGCGAACAGCACCCCGCGGGGGTCCGGC  
 CGCCAGCCGCCCGGGCGAGCAGCCCGCGCGCGCCTCGCGCGCCTCGGGGGTCCCGCGC  
 GCACCGCGCGCGCGGAGCACGGACTCCAGGACATCGGGCCGAGCAGCCCGCGAGCCCG  
 GTGGCCAGACGCGCGCTGCTCGGGGACGACGGGGTGGTTCAGCTCCAGGTGATCCGG  
 CTTCCGGCGGGTCCGAGAGCGCGGGGCCGACGCCCGCGCGCGCGCGGACATAGGTG  
 30 CCGCGCCCCACCTCGCGGACGGTGAGCCCTCTGCGGGCCAGCTCCCGGTAGACCCGGGCG  
 GCGGTGGAGTCCGCGATGCCGACCCCGGGGCGAACTCCCGTTCGCGCGGAAGCCGGTCC  
 CCGGGGCGCAGCCCGCCCGTCTCTGATCTCTCGCGACCGCGTCCGGCCACTGCCGGTAG  
 TCCTTCATCTCCCGTACCTCCCTGTCCGGTGGACCGCTTCCCGCCCCGCCCCGCGGACC  
 GTGAAACGGGAAGCACCCGTTCCGGAGTCCGAGCTCCCGTCCCGGAAGCTCCCGTCCGG  
 35 AAGCTCCCGTTCCAGAATTGCACCGAGAGCAATATTCCTATTGCACCGATCAAAACAC  
 CGATCTACGCTCGGAATTGCCTCACACAGACCGTCCAGCATCTGCCGCACACCGGTACT  
 GACGCCCCGTCCGACCGCACCCGCGCGGAGCCGTCCGCCCGCGCGCCCCGTTCGCGCACA  
 GGAGAGAGAAGGAGATGGTGGAGACCAGCGCACTCGCCGGTGTGGTGTGGTTCGCCCTCG  
 GAATGGTCTCACCCCGGACCGAACATGATCTATCTCGTCTCCCGCAGCATACCCAGG  
 40 GCGGACGTGCGGGGATCATCTCGTGGCGGTGTGGCCCTCGGTTTTCTGCTATCTGC  
 TCGCCGCGAATCTCGGCTGTGGTGATCTCTCGTCCCGTCCCGGAGTTGTATGTCCGG  
 TCAAATGGCCGTGCGGCTATCTGGCATATCTCGCTGGAACGCCCTGCCGCCCGGTG  
 GCGTGAATGTGTTCTCCCCGAGGAGGTTCGCGACGACTCCCCGAGCAGGCTGTTACCA  
 TGGGGCTGATGACGAACATCCTCAACCCCAAGATCGCCGTATGATCTCGCACTCATCC  
 45 CGAGTTTCTCGCGACCGAACCGGACCGTGTCTGTTCCAGGGGCTGATCTCGCGGTC  
 TCCAGATCGCGGTGAGCGTCCGCGTCAATCTCGCATCTGCTGGCGGCGGAGCCATCG  
 CCGCCTTTCTCGGCGGCCACCCCTTCTGGCTCAGGGTTACGCGCCGCGTATGGCGCGG  
 CGCTCGGTACGCTCGCGGTCTCCCTGGCCCTCGACACCTCCGCCCGCGCGCACCCGTCT  
 CCTGAGGCCCGCGGACCGGGAGCCGACGCGAAGGCACCCCTGGGCAACCGTTCCGAGAGC  
 50 TTATCCGTTACCCCATGAATCCGATATAAGTGCATTGGCCACTTACCATGCATGGAAC  
 AGGCCAACCTGACCAAAAAATGAGCCCTCCCCACCCGGAATAGATGCTTCCAGTGTGAA  
 GAAATTTATAGCGGGAGCGTCTGCCGAACAGGACGGCCATACGCCGCAAGGCAGAAGC  
 GACATCGCCGCCCGCCCGGTCCAGAAAATTCCGAGGACACATCGGACGACCGTCTCCGC  
 ATCGCGCTCAACTCCCGATTACAGAGAATATTGAGTACGTATCAACCGGGCCTTGATCTA  
 55 CTCAGCTCCATTGTTCTCTCCAGTCGGGATGTGCAATGAAGTACGACATAACCCACCA  
 TCCGGCTTCCGTTCCGACTCTCGGCCCGTTGACCGTGACCGCGCGGAGCAACCCGTG  
 GACCTGGGCGCGCCACGGCAGCGCGCCCTGCTCGCCCTGCTGCTCATCGATGTCCGCAAC  
 GTGGTCCCGCTGCCGGTCTGACCGCGTCTGATCTGGGGGGCCGACCCACCGTCCCGGGT  
 CGGGGGACGCTCCAGGCTTATGTGTCCCGACTCGCGAACTCTGCACCGCCATGACCGT  
 60 TCCCTTCGCTTGTCCACAGCTCCAGGGGTATCTCTCGAAGTGGATTCCGGCGAAGGTG  
 GACGCCGTGGTTTTTCGAGACACGTGTCAAGGAGTGCCGGGAATTGAGCAGGGCCCCGAAC  
 CCGGAGGCCACCCGGGCGTGGCTGGTCCGCCCTGGAGATGTGGAAGGGCACACCCATG  
 GGCGAGTGCATGATTATGAATTTGTGGCGCGGAGGCCGACCGGTGGAAGGAATCCGG  
 TTACGCGCGCTGGAGACCTGGTCCAGGCGTGTCTCGATCTCCAGCACTATGAAGAGGTT  
 65 GCATTTTCAGCTCGGCGAGGAGATCCACCGCAATCCGGAACCTGGAACGGCTGGGCGGTCT  
 TTATGCGGGGCCAGTATCATTCGGGACGGTCCGGCGGAAGCCCTGTTGACGTATGAACGT  
 ATGCGTACCGCGGTGGCGGAGAATCTGGGGGCCGATATCAGTCCGGAGCTCCAGGAACCTC  
 CATGGAAGATTCTGCGCCAGGAACCTCACGGAGACACCCCGCGCGATCGACGGCCTCC

CTCACACGGGCGGCGGGCCGCGACGGGCCCCGCCCTGGCCGAAACCGGCACCCCGCC  
 GCACCCGCGGACATGGCCGAAACACGGTGGCGGAGGAAAGCGCCGCGCCCCCGCCCCG  
 GCGGCGCCCGGACCCCGCCCCCATGCCGTCCCCGTACCGCTCCCCCATCCGTACGGG  
 5 GCGGTCCCGCCGGTCAACCCGGTGCCTCCCCCGTCCCCCGCTCGGCCCTCCGTTCAGCG  
 GCACCCGCGGAGACCGAGGACCCGGAACCGGCGCCGCCCTCCCCCTCCGCGGGGCGGC  
 CGACTCATCGGCCCGCGCCGAACTGCGCAGGCTGCGGCTGCTGCTGACGAAGACCCGC  
 GCGGGCCACGGCATGTCTGCTGCTGCGGCGAACAGGGCATCGGGAAGACCCGGCTC  
 CTGGAGCACACCGAGCACACCTGGCCGCGGGCGCGTTCCGGGTGGTCCGTTCGCACTGC  
 10 GTCGCCACCTCCCGGCACCGGGCTACTGGCCCTGGGAGCACCTCGTACGCCAGCTCGAC  
 CCGGACAGCGGCCTCGGTGACGACGGCGACGCCGACCCCGTCCGCCAGCGCGAGTGGCTG  
 CCGGAACACCACTTACCCACAGATGCGGATCTGCCGGACGGTGTCTGCCGCGGGCGCG  
 CGGACCCGGTCTCTGATCTCTGGAGGATCTGCACCTCGCCACGCGCGGTCTGGAT  
 GTGCTCCAGTCTCTGGTCAAACAGATCGGCCAGGCCCGCTCATGGTCTGCGCCACCTG  
 CGCGAGCACGATCTCGCCGGGACCCGCGCTCCGCCGGGCGGTGGGCCGCATCTCCAG  
 15 GCGGGCAACACCGGCACCCCTCCGGTGGACGGGCTCACCGAGGAGCAGACCGGGAGCTG  
 ATCGTCTCGGTGCGGGGGGCCCGTTCCGCGCCCATGACGCCAACGGCTCCAGCGCGCC  
 TCGGGCGGCAACCCGTTTCTGCTGCTCAGCATGGTACAGGGGAGGACGGCACCCAGGAG  
 TGGGCACGGCGGTGCGTCCCGTTGAGGTGCGCGAGGTGCTGCACGAGCGGTGAGCGAA  
 TGCTCCCCGTCCACCCAGGACGTGCTCAGCTCTGCGCCGTGCTCGGCATGAGCGTGC  
 20 CGACCGCTGCTACCGACATCATGTCCACGCTCGACATCCCGCACACCGCGCTCGACGAC  
 GCGCTCGGCACGGGCTGCTGCGCCACGACCGGAACACCGACGGAATGGTCCACTTCGCC  
 CATGGGCTGACCCGGGACTTCTGCTCGACGACACCCGCGGTACCCGCGCCCGCTGG  
 CACCACCGGTGCGCCGACCCCTCGCCCTGCGCTTCCAGCAGGGCGACGACACGCGGAG  
 ATCCGCGCCACTGTCTGGCCGCGGCCGTCTGCTCGGCGCCCGCGGGGGTGGCCCC  
 25 CTGCTGGCGTGGCCGACCGGGAGCAGTCCCGCTTCTCCACGCGGAGGCGTGGCGTGG  
 CTGGAGAGCGCGTTCGCGGTGCTCGCGCGCTGCCCGGGACCGCGGTGTCGCGCGTC  
 GAACTCCAGTTGCGCAAACGGATGATGGCGCTGCACGCGCTGATGGACGGCTATGGATCG  
 GCGCGCTCGAGACGTTCTCTCCAGGTACCCAGTGGGAACACGTCTTCGACAACACC  
 CAGCCCACCGGGTGTCTGCACGTCCAGGCGCTGAGCGCGCTCACACGGGCGCCATGAG  
 30 CAGGCGGCGGAGTGGCCGGGTGCTGCACGAGTGGCCGACACGGCGGCGGACCGGAG  
 GCCCGGTGCGCGCGCTGCTATGTGGACGGCGTCAACCTGTATGTGGGCGGACGGGTGAC  
 GAAGCCCTCGCCGCGCTCGCCAGGGCACCGAGATCACGGACGCCCTCTGGCCGGACAC  
 CGCAGACCGCCGCCCGCACGGCGGGGACCTCCAGGACCGCGCTATCGACTTCGCG  
 GCCTATCTGGCGTCCGCGCACTGTCTCAGCGCGACCGGATTCAGACCCAGCGCTACCGG  
 35 ACGGAACCTCTCCACTCACCCAGTCCGGAACGGTACGACCGGCCGTGGGACCGGGCCTTC  
 GCCCGCTATGTGGACGCGCTCATCGCCGTACCGAGTGCATGTCCAGGGGTGTGGCTG  
 GCCGCGCGGGCGGGGTGACCTCGCCGCCCGCTGCCAGTCCCGTTCTGGCAGCGGATG  
 CTCGCGTCCCCCTCGGTGGGCCGAGTCCACCAGGGGGCGCACGACAAGGGGTGGCC  
 CGGATGCGGGAGGCGTGCACGAGGCGGGCCCGGCACCGACCTGCTGCGCGTACGCTC  
 40 CACCTCGGCCTGCTCGCCGACGCCCTCCAGTACACGGGCGCCCGGGAACAGGCCCGGCGC  
 ACGATGTCTCCGCGTACGGGAGATCGAGCGCGCGGCGAGTACTTCTGTCTCCGGCCG  
 CAGTGGCCCTGGGCGCGGTCTCTCCACAGCCACGGCACCTCCGCCGCGGCGGAGACCGG  
 GTCGTCCACGGCAGGCACTGACCCGGGGCGGCGGAGCCGGGCCGTACGGTACGGGTG  
 CGGTCCGGGAGGCGTGCACGAGGCGGGCGGGGCGGGGCGGCCGACGGTTCCGGGGC  
 45 CGGCGGTGTGGGAGGGGGCGGCCCGGATCGCTCAGACCGGGCAGACGGCGGACCGCG  
 CCGCGCCCGGCGGAGCGCGGCCCGGCGGCGGAGTGGCGTAGTCCGCCCGCAGGAAGAC  
 CAGGGGCGAACCCTCGCGGATACCCCGAGGTGCGCGACCGCCCGGTGACGAACAGTG  
 GTCGCGCGCCTCCGTCTCCCCGCCACCTCGCAGTCAACACCGCGAGCGCGTTCGAGCAG  
 50 GACGGGGAGCGGTGGCGGTGCTCCGTACGGCACCTCCAGCGCCCCGATCGCCCCC  
 GGCGAAACTCCGGCAGACCGGGCCCTGATCCGCGCGAGCACATTGACGAGAAACGCCC  
 GGCGCGCGGAGCGCGGCCAGGTGCTGACGACCTGGCCGGGAGGAAACCCACGACAC  
 CGGATCGAGCGACACCGAGTGAACGTCCCCACCACCATGGCGGGCGGCGGTGCCCGG  
 AGCCTCGGCCGGACCGGTGACCAGGACACCCCGGTGGGATAGTGGCCCGCACCCGGCG  
 55 CAGCAGACTCCCGGACACGGACCCGTGGGTGTGCGCGGAAGGCCCGGAGCGCGGTAC  
 AGCCACGGTAACGCGCGGTGTCTTGGCCGCGTAATCGGGGTCCAGATAGACGAAGGCC  
 CGGTGGACGAGGAAGTCCCGCACCTCGTAGACCGTGCACAGCGCCCGGCGGCGCCACTCG  
 GGGTACCCGCGCGCACGCGCGTCCCGGTGCTACCGTGGGTGGTGCCTCCGCGGCG  
 AGGAGTTCGGTCCCGGTGAGAATCCAGTTGACGGACACAGATGGTGGGTGATCGAGCGG  
 60 ATGGTCCCCCGAGGTGTCGAAGAGCGGGCGATCTCGGACTTCCCCGGGCGCAGACCC  
 CACTTGGGGAAGAAGAAGACCGGTCTCTCGGCGAAGTAGTCGATCGGGGGGTGCCGTG  
 CTGCCGACGCCCGTGTGCGAACGCTTGAAGTACGCGGTGATGACCGCCTTGGCGTGC  
 TCGTCCGTATACCGGCCGATGCCACGGACATGAAACGACCTCCAGAGATTCCGGGTGGC  
 TGTGCTGGGCTGCGGAAGGGGTGTCGCCCGCGAAGGACGGCGGACGCCGCGGACGCCG  
 65 GCGCTCTCCCCGGGACGGGTCCAGCGTCTGGAGAGGGCTTGGCGGCGGCTTGACG  
 CCGTGTGTCGCGCGGTGCGGAACGCGAAGTACCGCCAGCTACGGGCGTTGCACCG  
 GACGTGTACCGGTCGGGACCCCTCGTACCCCGGAGCGCGGCGACCCCGGCGGCTCCG  
 GGGGTACGACGCGCGGACCGGCCGAGCGAGCCGACGGGTGCGACGGTGCAGGTGGT  
 TCCGGTGTGTCGACAGCTCGGACGGACCGGACGGTGCAGGTGGTTCGGTGTGTCGAC

AGCTCGGACGGGTGCGACGGTGCGCGTGGTTCCGGCACGCCGGACGGGTGAGTTGCCGAT  
 CATGGCGAGCAATGCCGGGTGTACCGCTCCCCGACACCGGGTGGGAGATCGCGGCCGT  
 CACCTCCGCGAGGGACCGGTGTCAGCCGGATCGAGGCGGCGGCGAGATTGTCCGCGAG  
 5 ATGGGCCGGGTTCGCGGTGCCGGGATCGGGACGACGTCTCGCCCCGGTGGTGCAGCCA  
 GGCGAGCGCGAGCTGTGCCAGGGTCAGCCCCAGACCGTCCGCGACCGGGCGCAGCCGGTG  
 CAGCAACGAGCGGTTGCGCGCGAGGGCCGGAGCGCTGAACCGGGCTGGCCCCGGCGGAA  
 GTCTCGTCCCCAGATCGTGGTGGTGGGATGGTGCCGGTGAGAAAACCCCGTCCCAG  
 AGGGGCGTAAGCGACGATCCCGATCCCCAGCTCCCGGCAGACGGGCACCACCTCGTCCTC  
 10 GATCCCGCGCGACACAGGCTCCACTCGCTCTGCACCGCCGTACCCGGTGCACCGCGTC  
 CGCCCCGCGCAGCGTGGCCGCGAGGGCTCGGAGAGACCGAGCCTGCGGACCTTGCCCTC  
 GCGCACCAGCTCGGCCACCGCACCCACGGTCTCCTCGATCGGCACCGCCGGTCCGTCCA  
 GTGCTGGTAGTACAGGTGCGTGGTGGTGCCGAGACGACGAGGGACCGTTGCGAGGC  
 CGCGCGGACGTAGGACGGCTCGCCGCACAAGCCCTGGGAGGCGCGTGGACGAGCGCAC  
 15 CATGCCGAACCTTGGTGGCGATCAGCACCTCGTCCCGCGGCGCCGCGACCGCCCGTCCGAG  
 CAGCTCCTACCGGCGCGGAGCCCTGGACGTGCGCGGTGTCCAGCAGGGTGACCCCGGC  
 GTCGACGGCGCGCGGATGGTGGCCGTGCGCCGGGCGCGGTCCGGGCGTCCGTAGAAGTC  
 GGTGGTGGCAGGCAGCCGAGCCCTGGGCACTGACCGGAAGTCCCGCAGGGCGCGGAC  
 CGCGGACGCGGAACCGCGCGGACACGGAACCGGCCGGGACTCGGGCGGAGAGCGGGA  
 20 CATAAGGAACCTCCACAGGCGGAGCCGGGAACGGGACGAGGGCGAGGACGGGACGGAACG  
 AAGGAGAGGACGGGACGGACAGCACGGACGGGACGGACGGAACGGAGTCCGGAACCGGGG  
 GGGGTGACCGGAACCGGGCCGTCTTGGCCCTCCCCGTCCTCCCCGCCATCCGCCGTTT  
 TCCCCGTTCCCTCTCCCGTCTCCAGCCAACACCGCCGCCCTTTCCAAGCGCTTGACAC  
 GGCACCGACAGCCCGCGGGCGCCGATGGGGACCGTGCCCGCCGGTGAGCGGCGGT  
 25 GAGCGCCGGTACGGGACCCACGCGCCGCGCCCGGGCGCCCGCAGGGCCCGCGCGGCC  
 ACCCGGCGCCCGCCCGGCGGAGCGGCGATCCGGGCGCTCGCTGCAAGAGGAACATCCA  
 CAGCCGCACAAGGAGCGCTCCGCACAGTGGGCACCACGTCCGCCCCGTCCCCACACCGT  
 GGCCGGTCCCCACCGACAGCACAGCACCGCACAGCACCATCGCACGGCACAGCACAG  
 CACCACCGGCACGAGGAACCAAGGAAAGGAACCAACCAACCATGACCTAGTGGACTGCA  
 30 CCGCGTACGGCCCGAGCTGCGCGCGCTCGCCCGCCGGTGGCCCGGACCCCGGGCCG  
 ACCTGTACGCCCTTCTGGACGCCGCGCACACAGCCCGCCCTCGCTCCCCGGCGCCCTCG  
 CCACCGCGCTGGACACCTTCAACGCCGAGGGCAGCGAGGACGGCCATCTGTGCTGCGCG  
 GCCTCCCGGTGGAGCCGACGCCGACCTCCCCACCACCCCGAGCAGCACCCCGGCGCCCG  
 AGGACCGTCCCTGCTGACCATGGAGGCCATGCTCGGACTGGTGGGCGCCGGCTCGGT  
 35 TGACACAGGGGTACCGGGAGCTGCGCTCGGGCAGGTTTACCACGACGTGTACCCGTGCG  
 CCGGCGCGCACCACTGTCTCGGAGACCTCCGAGACGCTGCTGGAGTTCCACACGGAGA  
 TGGCTTACCACCGGCTCCAGCCGAACCTACGTGCTGCGCTGCTCCCGGGCCGACCAG  
 AGCGCACGGCGGCCACACTCGTCCGCTCGGTCCGCAAGGCGTGGCCCTGTGGACGAGA  
 GGACCCGGGCGCGGCTCTCGACCGGAGGATGCCCTGCTGCGTGGATGTGGCCTTCCGCG  
 40 GCGGGTGGACACCGGGCGCCATCGCCAGGTCAAACCGCTTACGGGGACGCGGACG  
 ATCCCTTCTCGGGTACGACCGCGAGCTGCTGGCGCGGAGGACCCCGGACAAAGGAGG  
 CCGTCCGCGCCCTGTCCAAGGCGCTCGACGAGGTACGGAGGCGGTGTATCTGGAGCCCC  
 GCATCTGCTGATCTCGACAACCTCCGCACCAACGACGCGCGGACGCCGTTCTCGCCCC  
 GCTGGGACGGGAAGGACCGCTGGCTGCACCGCTTACATCCGCACCGACCGCAATGGAC  
 45 AGCTCTCCGCGCGGAGCGCGCGGGCGACGTGCTGCGCTTACACCGCGCGGCTGAGTCT  
 CCGGGTCCGACACCGCGCGGCTGAACCCAGGTCCGGGGCCACGGTCCGGCACCGCGCG  
 GCTGAGCCCCCGGGTCCGGCAGCGGGCGGCTGAACCCCGCCCGGGCCACCGCCCGACC  
 GCCCCCGCGCACCGGACGCGCCCGCTGTACGGCGGTCCCGCCGGGCCCCGTACACCTGA  
 AGCGCCCGGCGGACCGCGCCCGCCCGGGGACGGACAGAGCCGGGTGCGGGAGGACGTC  
 50 CTCCCGCACCCGGCTCCACCGTTCCGCACCGACCGCACCCGACCGTCCCGCAGGCGCCA  
 CCGGCACCGCACCGCCCGCGCGCGGACGCAACACAGGCGCCACGCGCCCGCACGGTGCC  
 CGCGCTGCTCAGCCCCGTCCACCGGGCTGTCCAGCAGCCGCGCAGCGCGCCCCGATG  
 AACTCCCGGTGCGGCGCCGACCCCCCGGACCCCGCGAGATGCCCCACACTCCCGGGATC  
 ACCTCCAGCGAGGCATACGGCAGCAGATCGGCCACCCGTTCTCGTCTCGACGGCGAAA  
 55 CACACGTCCAGGGCGCCCGGACACACGGCCCGCGCGTGAAGGAGGCGAGCGCGCC  
 TCGACGCTCCCCCGGCCCGGGTGTGCCCCACATCCGTGTTCTCCAGGTGCGCACC  
 ATGGTGAGCAGATCCGCGCGCGCGGGCCCGGAGAGGAAGACCTGCTCCAGAAGCCGGTG  
 AGGTACTCTCGCGGTGGCGAAACCCAGTCCCGGTGGGCACGGCGGGCCAGAAGGAA  
 CGCGAGGTCCCCACCCGCGGAACACCCGGCCCGCGCTTCCGCCCCGCTCCCGGGCG  
 60 TCGGCGCTGAGCGCGCGGCGGACCGGACAGCAGGACAGGCTGTGCGGGTGTCTACC  
 GCGCCCCCGAGATCGGGCGATCCGGCGACCATCCCGGATGCGACACGGCCACTGG  
 TAGGCGTGGGCGCGCCATCGACAGCCCGTGACAGGGCCAGTTCCCGTACCCCCAGC  
 TCCTCGGTGAGCAGCCGGTGTGCGCCGCGACATTGTCTGCGGAGTGATCAGCGGAAAG  
 CGGGACCCCGACGGGTGGTGGCGGGCGAGCTGGAGACCCCGTTGCCGAAGAGTCCGGCG  
 65 GTGACGACGCGATACCGCCGGGTGTCCAGCGGCGAGCCCGCACCGATCAGCCAGTCTGAC  
 CCGGTGTGGTCCCGCGAAGAAGACGAGACAGACACCGTTCGTCCCGTGGCGGTTT  
 GGCGTGCCGTACATGGCGTAACCGATCCGGGCGTCCCGCAGGACCTCCCCGTCCAGCAAC  
 GGCAGTTCGTGATCTCGAATATGCGGCATTCCACCGCTGACCTCCTTGTTCGATCCCC  
 CGGACAACAGGTGCGTCTG GCCGGAGACTCAGAGCCAGTTGGGGCGGATCTCGGTGGCC



CACAGCTCCAGGCTGCGCAGCTGGACATCGTGCGGGATCAGCCCGAGTACTGGCACTGG  
 AGCAGATACTCCGGATCGTGCCGCTCCACCAGCTTCTCGATCATGCGGTGATGTCGTCC  
 GGGGTGCCGACCCACTCCAGCCCCGGTCGACCAGGCTCTGTAGTCCGAGCCGATCGGA  
 CCCGTCTCGCCGGTCGCGCGCAGCGCTCGGTGAAGCCATGGGGCCGAACCACTTCTCG  
 5 AAGATGAAGCCGCCCGCGCGGACGCCAGTGGTGGGCCTCGCCGGAGTCCCGGGAGACC  
 AGGACGTCCTTCATCACCCCGACCCGCTCGCCCCGCGCAGGGTGCCGTGGCCCGCGCC  
 TCGGCCTCCTCCCGGTAGATGTCATCAGCCGGGCGACGATCTGGTCTCGGTGTTTCATC  
 AGGATCGGCACCCACGCCCTCCCGGGCACAGAACCAGGACGTGTCTCACTGAAGCTGAAC  
 GGCTGGAAGACGGGCGGGTGGGGCGCTGGTAGGGCTTGGGCGCGATGCCACCTCGCGG  
 10 ATGACGCCGTTCTCGTCGAGGCCCGCGCGTAGCGGCGCACCGCCTCGTAGGGGAACCTC  
 AGGTCCGGACCCGGATCGTCCACTGCTCCCCGAGTGGGTGAACGTCCTCGGTCTCCAC  
 CGCTTCTTGATGATCTCCAGTGCTCTCGAAGAGGGCAGATTGCGCCGGTCCCGCTCC  
 CCGGCTCGGACAGGGTGCCGCCGACCCGTACACCTGCCCATGATGTCGGCCAGCGC  
 TTCTGGAACCCGCGCGGATCCCGACGAAGGCGCGGCCCGGGTCATGTGGTCGAGCATC  
 15 GCCAGATCTCGGCCAGCCGCGAGCGGATTGTGACGCGGACGAGCTTGGCCATCTGGCCG  
 ACCCGGATGTGCCGGTCTGCATGCCGAGGTAGAGCCCGAGCATGATCGGGTGTGTGGAG  
 ACCTCGAAACCTCGGTGTGGAAGTGGTGCTCGGTGAAGGACAGTCCCCAGTAGCCGAGT  
 TCGTCGGCCGCTGCGCCTGCGGGTGAGCTGCCGAGCATGTTCTGGTAGTCTGTCGGA  
 TTGACCCCGCCATACCCCGCTGGACCTGCGCATGACTGCCGACCGTTGGCAGATAGAAG  
 20 AGAATGGACTTCACCTGGCTCCTCCGGTTCGCGGCGCCCTCCATTGACGTGCGCCGAAA  
 GCGGCTCGACCGTCCCACTCCGCCCTTGAGTTCCGTCTGACGCCGCGCCAGTCCGCGGGC  
 CGTCCCGCGGGGTGCCCGCGGGGTCCGACCCGCGGACGGCACGGCGCGCACCGCGCG  
 CGCGGCGCTTCGGGACCGGGCTCGACGGGTGCTCAGCGGGACGTCCAACGGAAGGCA  
 AGCCCCCGCTACCCAGCTTGGTCAAGGCGCTCATCGCCATTCCTGAGGAGTCCCGCCTT  
 25 GACCACAGCAATCTCCGCGCTCCCGACCGTGCCCGCTCCGACTCGAAGCACTGGACCG  
 TGCCACCCCTCATCCACCCACCCCTCTCCGAAACACCGCGGAACGGATCGTGCTGACCTC  
 GGGGTCCGGCAGCCGGTCCCGGACACCGACGGCCGGGAGTACCTGGACGCGAGCGCGT  
 CCTCGGGGTGACCCAGGTGGGCCACGGCCGGGCGGAGTGGCCCGGTTCGCGCCGAGCA  
 GATGCGTCCGCTGGAGTACTTCCACACCTGGGGGACGATCAGCAACGACCGGGCGGTGGA  
 30 GCTGGCGGCACGGCTGGTGGGGCTGAGCCCGGAGCCGCTGACCCGCGTCTACTTCACCAG  
 CGGCGGGGCGGAGGGCAACGAGATCGCCCTGCGGATGGCCCGGTCTACCACCACCGGCG  
 CGGGGAGTCCGCCCGTACCTGGATACTCTCCCGCGGTTCGGCCTACACCGGCTCGGATA  
 CGGACGCGGCGGCTCACCGGCTTCCCGCCTACCACAGGGCTTCGGCCCCCTCCCTCCC  
 GGACGTGCACTTCTGACCCCGCGCAGCCCTACCGCCGGGAGCTGTTCCCGGTTCCGA  
 35 CGTACCGCTTCTGCTCGCCGAACGCGCGAGACCATCGACCGGATCGGCCGGAGCG  
 GATCGCGCGGATGATCGGCGAGCCGATCATGGGCGCGGTTCGGCGCGCGGCCCGCCGCG  
 CGACTACTGGCCCCGGGTTCGCCGAGCTGCTGCACTCCTACGGCATCTGCTGATCTCGGA  
 CGAGGTGATCACGGGGTACGGGCGCACCGGGCACTGGTTCGCCCGCGACCACTTCGGCGT  
 40 GGTCCCGGACATCATGGTCACCGCCAAGGGCATCACCTCGGGTATGTGCCGACGCGCG  
 CGTCTCGCCGACCGAGGCGGTTCGCCGAGAGTTCGTTCGGCGACCGGGCTTCCCGCGGG  
 CTTACCTACAGCGGCCATGCCACGGCTGCGCGGTGGCCCTGGCCAACCTGGACATCAT  
 CGAGCGCGAGAATCTGCTGACAACGCCAGCACCGTTCGGCGCCTACCTGGGCAACCGCT  
 GGCCGAGCTGAGCGATCTGCCGATCGTCGGGACGTTCGGCAGACCGGTCTGATGCTCGG  
 45 TGTCGAACTGGTTCGCCGACCGCGGAACCCGGGAGCCGCTGCCGGGCGCGCGCTCGCCGA  
 GGCCCTCGCGGACCGGGCGGCGATCTGCTGCGCGCCAACGGCAACGCCCTCATCTCAA  
 CCCCCGCTGATCTTACCCAGGAAGACGCCGACGAACTCGTGGCGGGCTGCGCTCCGT  
 ACTCGCCGACACAGGCCGAGCGGCGGGTGTCTGACCCCTTTGGCCCTCCCGGCCCG  
 ACCGGGGACACCCCGCGCGACCCCGAGCGCAAAAGACCCCTCTGCCTGCGTTTCCGC  
 AGGTCAAGGGGTCTGGTGCAGTGGAGCCTAGGGGAGTGAACCCCTGACATCTGCCATG  
 50 CAAAGACAGCGCTTACCAACTGAGCTAAGGCCCGAAGCGACAGAACGGCCCTGGACTG  
 CTCCGTCCCGGCCACTGCCGAGACAGAGTACCGGGTGTTCGGGTGATCTCCAAAC  
 ATTGAGGTCTCCCGGTGGGCGACCACTCTCCGTAAGATGCTCGACGTGGTTTCGACGAGC  
 GAAGCCCGCTTGGGGAAGCGATGGGGAGACGCGCATGGACGCCGCTCAGCAGGAGACGAC  
 CGCAAGAGCCCGGAGCTACAGCGAAGCTGGTACGGGGAGCCCTGGGGGCCCTGTTCCG  
 55 CAGGCTGATAGCAGTCTGGGGCTGAACAGGCGCGTCTCGCGCGGTGCTGGGCCTCTC  
 CGCCCCCATGCTCTCCAGCTCATGAGCGGCCAGCGGGCCAAGATCGGCAACCGGCCGT  
 GGTCCAACGGGTCCAGGCGCTCCAGGAGTTGGCCGACAGGTGGCCGACGGCAGCGTCAG  
 CGCGGTGGAGGCCACCGACCGCATGGAGGAGATCAAGAAGTCGAGGGAGGCTCCGTCTC  
 GACCGGAACAGCCAGACCAACAGCTCGGGGGCGCGACCGTCCCGCGGGTCTGTCG  
 60 GGAGATCCAGTCTGCTGCGGTCCGTGTCCGCCGCGGGGACATCATCGACGCGGCGAA  
 CTCCCTCGCCCCGACCATCCGGAGCTGGCAGAGTTCTGCGGGTGTACGGGGCCGGGCG  
 CACCGCGGACGCGGTGGCGCACTACGAGTCCCACAGAGCTGACGACCGAGGCGCGCCCC  
 GGAACGGACAGAGCCTCATGAGGACGGGGAGCGGACGCGGCACCATGGGTGAGGTCTT  
 CGCCGGCCGCTACGAGTGGTTCGACCCGATCGGACGCGGAGGGGTTCGGCGGGTCTGGCG  
 65 CGCTGGGACACCGGCGCGCGCTATGTGGCGGCAAGGTGCTCCAGCAGAGCGACGCG  
 GCACACCTGCTGCGCTTCGTCCGCGAGCAGGCGCTTCGGATCGACCATCCCATGTCCT  
 GGCCCGCGGAGCTGGGCGCGGACGACGACAAAGTCTTACCATGGATCTCGTGGG  
 CGGCGGATCACTCGCGACGTGATCGGCGACTACGGCCGCTCCCGCGCGCTATGTGTG

CGCCCTGCTGGACCAACTCCTCTCCGGGCTCGCCGCGGTGCACGCCGAGGGCGTGGTGCA  
 CCGCGACATCAAACCGGGCAACATCCTGATGGAGGCCACCGGGACGGGCGCCCCATCT  
 GCGCCTGTCCGACTTCGGCATCTCCATGCGCAAGGGCGAGCCCCGGCTGACCGAGACCAA  
 CTATGTCTGTGGGTACGCCCCGGTTACTTCGCCCCCGAGCAGGTGAGGGCGCGGAGCCGGA  
 5 CTTCCCCGCGCATCTCTTCGCCGTGCGCCTGGTTCGCCCTCTATCTGCTGGAGGGTCAGAA  
 ACCCGACACCAAGGCCCTGGTGGACTTCTTCACCGCCCATGGCACCCCCGGTGTCTCCCG  
 GGGGATACCGGAGCCGCTGTGGCAGGTGCTCGCGGGGTGATCCAGCCCGACCCCGCCG  
 CCGGTTCCGTACGCGACGGGGGCCGGAAGGCCCTCGCCGCGCGGTGGAAGTGTCTCC  
 CGAGAGCGGCCCCGACGACGAACCGGTGGAGATATTGACCAACTGGGCCCGCTGCCGCC  
 10 GGGGTTTCGGCCCCGCGGCCCGGAGAACACGCGGCCCTCCGGTCTGCTGCGCTCGGCGGC  
 CTCCGGTACC

# SEQ ID NO:18 orf2par reverse complement

ATGGCCACCACGACCGGAAAGCCATGCTGGAACGTTTACCAGTACGGTGTGACCATGTATTTCGGCGTCTGCG  
 15 GCCGGGAGGCGTCCGCCATTCTCTTCGACGAGGTGGAAGGACTCGACTTCGTCTTGACCCGGCAGAGTTTACC  
 CGGGGTGATGGCGGACGTCTCGCCCGGATACCAACCGCCCCAGGCGTGTTCGCGACCTGGGCCCCGGCATG  
 ACCAACCTGGCCACCGCGTTCGCCACCTCCGCCCTGGACCGCAGCTCGGTTCATCGCGTGGCCGCGCAGTCCGAGT  
 CGTACGACTGCTACCCCAACGTACCCACCAAGTGCCTGGACAGCACCGCCGTGATGGGCCGCTGACCAAGTTTACG  
 CGTCCAGCTCGAACCGCGGCGAGGACATCGTCAACCTCGTCGACAGCGCCGTCTCAACAGCCGGATCGAGCCCGTG  
 20 GGTCCAGCTTCATCAGCTGCGGTCGACCTCGCGGCTCGAGCTGAACGGCACCCCGACCGAGCCCCCTGG  
 TCCGGGCCACCGCCACCCACGCCCTGGACGCGGACTGGCGCGCCCGCTCGACGAGGCGCTGAGCTGGTGGCGGA  
 GGCCGAGAACCCCTCTCGTCTGTCGTTAGCGCCGTTCATCGCGCGGGGCGGTGACGCGCTGCGCGCCCTCGCC  
 GAGCGGCTGAACATCCCGTCTGTCACCACTACACCGCAAGGGCGTCTGCGCGACGACACCCGCTCAACTACG  
 CGCGCATCAGCGGCTACATGGACGGCATTCTCGGCCACCCGGCCCTCGACGAGATCTTCGGCCCCCGGACCTCT  
 25 CCTGGCGATCGGCTACGACTACGCGGAGGACTCGCCCTCCAGTGTGGACGCGGGGCCGGCCAGACCAAGTCC  
 CGGGTTCGCCCCGAGGTCAACCCGATCCCGGAGTGTTCGCGCGGACATCGACATCGTACCAACGTTCGCCGAAT  
 TCGTACCCGCGTTCGACGACGCGACCTCGGGCCTCGCCCCAAGACCCGGCAGACCTCAGCGCCCTGCGCGCCG  
 CGTTCGCCGAATTCCTCGCCGACCCACCGAGTACGAGGACGGCATGCGGGTCCACAGGTGATCGACTGCATGAAC  
 TCCGTCTCGACAACGGCACCTTCGTGAGCGACATCGGCTTCTTCGCCACTACGGCGTGTCTTCGCCAAGTCCG  
 30 ACCAGCGTACGGATTCTCCTACCTCGCGGGCTGCTCCAGTTCGGCTACGGACTGCCCGCCGATGGCCGCGCA  
 GATCGCCCGGCCGCGGCGGCGTCTTCTCATCGCGGGCGACGGCGGCTTCCACTCCAACAGCGCGGACATCGAG  
 ACGGCCGTGCGCTGGGCTGCGGATCGTTCATGGTCTGTCGTCACAACGACCGCAACGGCCTGATCGAGCTGTACC  
 AGAACCTCGGACACCGAGCTCTCCACGCCCCCGCGTTCGGTTCGGAAGCGTTCGACTTCGTCCAGTTCGCGGAGG  
 CAACGGCTGCGAGGCGTTCGCGCCACCGACCGACCTCGTGTGCGCGCCCTCACCAAGGGCGCGGACTCGGC  
 35 CGCCCGTTCTGATCGAGGTACCGTGGCCTACGACTTCAGTTCGGCGGTTTCGCGGCCCTGGCCATCTGA

# SEQ ID NO:19 orf3par reverse complement

ATGCCCGGCCCCGACCTCGTGTACGGATTCCGGGTGCGCATCGGCACCGAGGGCGCCCCGGCGGGCGCCCCGGCG  
 40 GTCACTCCGAACCCGGCAGCGCACCCCGCTTCGCCGTTCGCGGGACCATGTCCCGTGCACGACGGCACCGCGTA  
 CCGCTCTGGAGCGGAACGGCCGTGACCTGGGCCGTTCGCCCGTCTGGTTCGCGACGGCCAGGTCCGGCTGCTC  
 CTGGCGGGCGAGCTGTACAACCGCGCCGAGCTGACCGGAGCGCTCGGCGGCTCCTCTGCCGCCCTCGGCGACGCG  
 AACTGCTGCTGGCCGCTTGGCGGCTGGGGCCCCGGGGCTTCGGCTCCTGAACGGACGGTTCGCGCGCACTGCT  
 CACCGACGCTCCACCGCGCGACCGTTCGCGGCCACCGACCGCGGTTTCGGTACCGCTGTGGTTCGCGCGCGGAC  
 45 GTGACGGGGCTGAGCGCCGCCACCGAGGCGAAGACCTGGCGCAGGACCGGGCGCGCGCTGGGCTGTCCGGCA  
 CCCACACCCCGCGGGGGCGCGGGCGTTCGCCGGTTCGCCCGCGGGACCGCCCTCCTGCTGCACGGAGTTCGCGGCG  
 TCCGACATACCGCCAGGGCGGTTCGACCTGGACACCCCGCTCTCCCGGGCGTTCGCCGGCGAACGGGAGGCGG  
 TGGACTGCTGGCGGAACGCTTCGCGACGGCGGTTCGCGCGCGGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG  
 CCGTTCGCGCGGCGATCGACTTCGGGGGAGTTCGCGGCCACACGGCGGCCCTGGCACCCGGGACAGGTTCGCTGTCG  
 50 ATGGGACCGAGGTGTCCGACGAGTTCGACGCGGCGCGCTCGGTTCGCCGTTCACCTGGGACCGCGCACAGCGAGA  
 TCCGGCTCCACTCGGCGGAACCTCGTCAGGGAACGCTTCGGGCGGTTCGCCCGCGCGGAGATCACCGACCCACGGT  
 CCTGGAGTACCTGCTGCCGTCTGTCGCCCTTACCGGCGGCTCGACACCGGGCGGCTTCGCATCTCACCGGGTAC  
 GCGCGGACATCCCGCTCGGCGGTATGCACCGCGCACGGCTCGCTCTGGTCCCTCGACGACGAGATCGCGGGCG  
 ACATGGCGGGCTTCGACGGCTCAACGAGATGTCCCGCTCCTCGCGGGCATCGCGGGAAGTGGACACCCACCC  
 GTACTGGGACCGCGCGGTCTGACGCGGTGGTCTCCCTCGAACCCGGGCTCAAACGCCGGCGGGGACCGACAAG  
 55 TGGGTGTTCGGGACGGCCCTCTCGGCGGTGTCGCCCGGAGACCGTGGCCCGCCCCAAGTGGGCATCCAGGAGG  
 GGTCCGGCACACAGCGGTGGACCGGACTGCTCCTCGCGGAAGGGATCCGGCGCGACGAGGTGACGGCGGTCAA  
 GGGCGCATGGCACGGCGCTGTACGACGCGGTGGTTCATCGACAGGTGCGCGCGGAGGAGTGGACTTCGGCGAG  
 ACGGTGCGGCGCTCCGTGACGCGGTGCGCAGGCTCAGGCTCAGGGCGGGTGGTGGTATGA

# SEQ ID NO:20 orf4par reverse complement

GTGTCCACCGCGTCTCCCCGCGTACGCCCAACGGGCGACCTTCATGCGGCTGCGCCACGGGCCGACCCGATCG  
 60 GCCATGACGTGGTGGTCTGCGCGCCCCGTACGACGAGGACACAGCTACCGGCCCGCGCGCGGTTTCGCGCGCGG  
 CGCCATCCGGCAGAGTCCAGCTGATCCACGGCGTTCGGATCGACCGGGGCCAGGGGTCTTCGACCGGATCGAC  
 GTGGTTCGACGGGGCGACATCGACCTCAGCCCTTCTCGATGGACCTGGCGATGGACACCGCGACGGTTCGCGCTGA  
 65 CCGGCTCCTGGAACGAACGACGCGTTCCTGATGCTGGGCGGGGACCACTCGCTCTCCCTGGCGGCCCTGCGCGC  
 CGTGCACGCCCCCACGGCCGGGTGCGCGTCTGACCTGGACGCGCACAGCGACACCAACCCACCGTCTACGGC  
 GGCACCTACCAACAGGACACCCCTTCCGCTGGGCCATCGAAGAGGGCCTGGTGGACCCGGAGCGCTGGTCCAGG

TCGGCATCCGCGGCCACAATCCGCGGCCCGACTCCCTGGACTACGCGCGCGGGCACGGCGTCAGCATCGTCACCGC  
CGCCGACTTCACCCGGCGCTCACCGCGCGGCATCGCCGAGCAGATCCGCGCGCACCGTCGGCGGCCTGCCGCTGTAC  
GTCTCCGTCGACATCGACGTCGTCGACCCGGCGTACGCCCCGGGCACCGGCACACCGCCCCCGCGGGCTGTCTT  
CGCGCGAGGTGCTGACCCTGCTCGACGTGGTTCGGGAGCTCAGGCCCGTCGGCTTCGACGTGGTTCGAGGTGTCCCC  
5 GCGGTACGACCCGTCGGGGATCACTCCCTGCTGGCGCGGAGATCGGGGCCGAAGTCTCTACAGTACGCCCCG  
GCCACCACGTCGCCCCGCTCGGCACCGGTGGACTCTCCCTGCCACCGGGGGCGGGCGGACGACGCCGAGAACG  
CCGAGAACCGCGGTGGACGCGGTGGACGCCGAGAGCGCGTGGACTTCGCCGGGCAGCGGTGGGGGTAG

## SEQ ID NO:21 cvm6 Polypeptide

10 VPGSGLEALDRATLIHPTLSGNTAERIVLTSGSGSRVRDTDGREYLDASAVLGVTQVGHGRAELARVAEQMARLEY  
FHTWGTISNDRAVELAARLVGLSPEPLTRVYFTSGGAEAGNEIALRMARLYHHRGESARTWILSRRSAYHGVGYGSG  
GVTGFPAYHQGFPGPSLPDVFLLTPQPYRRELFAGSDVTFCLAELETIDRIGPERIAAMIGEPIMGAVGAAAPPA  
DYWPRVAELLHSYGILLISDEVITGYGRTGHWFAADHFGVVPDIMVTAKGITSYVPHGAVLTTEAVADEVVGDQGF  
15 PAGFTYSGHATACAVLANLDIIERENLLDNASTVGAYLGKRLAELSDDLPIVGDVDRQTGLMLGVELVARGTREPLPG  
AAVAEALRERAGILLRANGNALIVNPLIFTQEDADELVAGLRSLARTRPDGRVL

## SEQ ID NO:22 cvm3 Polypeptide

20 VTRPPGLSAHTGVSVSGSLLRRVAGHYPTGVVLTGPAAEPGQPPPMVVGTFSTSVSLDPVLVGFLLPARSSTTWPR  
LRAAGRFCVNVLGADQGPVCRSFAGGDPGRWEVPYRTTATGSPVLLDALAWFDCEVAGETEAGDHWFTVGAVRDLG  
VIREGSPLVFLRGDYGHWAGGGSGRAGRRAVCPV

## SEQ ID NO:23 orf6par Polypeptide

25 MRASSPRGFRVHHGAGIRGSHADLAVIASDVPAAVGAVETRSRFAAPSVLLSRDAVADGIARGVVVLSGNANAGT  
GPRGYEDAAEVRHLVAGIVDCDERDVLIASTGPVGERYPMSRVRAHLRAVRGFLPGADFDGAAAVALGTAGARPTI  
RRARCGDATLIGVAKPGTGPAEQDDRSTLAFFCTDAQVSPVLLDDIFRRVADRAFHGLGFGADASTGDTAAVLAN  
GLAGRVLDLVAFEQVLGALALDLVRDVVRDSCGGALVTVRVTAHDEQAGRVGRAVVDAPSLRAAVHGPAPDWAP  
VAAVAGGHGDEGPGRSPGRITIRVGGREVFPAPDRARPDAVTAYPHGGEVTVHIDLGPGRAPGAFTVHGCDLLA  
GYPRLAGRAV.

## SEQ ID NO:24 orf4par Polypeptide

30 VSTAVSPRYAQPATFMRLRHRPDPIGHDVVVVGAPYDGGTSYRPGARFAPRAIRHESSLIHGVGIDRGPVGFDRID  
VVDGGDIDLSPFMDLAMDTATVALTRLLERNDATFLMLGGDHSLSLAALRAVHARHGRVAVLHLDAHSDTNPPVYG  
GTYHHGTFFRWAIEEGLVDPERLVQVGIRGHNPRPDSLDYARGHGVSIPTAADFTRRSPRGIAEQIRRTVGGGLPLY  
35 VSVDDIDVVDPAYAPGTGTAPGGLSSREVLTLLDVVGQLRPVGFVVEVSPAYDPSGITSLAAEIGAELLYQYAR  
ATTSPASAPVDSPLPPGAAADDAENAENAVDAVDAESAVIDFAGQRWG.

## SEQ ID NO:25 orf3par Polypeptide

40 MPGPDLVYGFVRIGTEGRPGGGPGGHSEPGSAPRFAVRGTHVPVHDGTAYPLWSGTAVTLGRPPVLVADGQVRL  
LAGELYNRAELTGALGGSSAALGDAELLLAAWRRWGPGAERLLNNGRFAALLTDASTGATVAATDHAGSVPLWLRAD  
VTGLSAATEAKTLAHEPGRPLGLSGHTAPGAAGVCRVPAGTALLHGVGGS DITARAVRTWTPPLSRALPGEREA  
VDLVGERLATAVRTRLRGGEAAPTIVLSGGIDSGGVAHAALAPGTRSVSMGTEVSDEFDAARSAVHLGTAHSE  
IRLHSAELVRELPAWAVAAAEITDPTVLEYLLPLVALYRRLDTGPLRILTGYGADIPLGGMHRRTASLWSLDDEIAG  
DMAGFDGLNEMSPVLGAGIAGKWTTHPYWDRAVLDAVLSLEPGLKRRRGTDKWLRLQALSGLLPAETVARPKLGIHE  
45 GSGTTSAWTGLLLAEGIRREDEVTA VKGAMARRLYDAVVIDTVPPEDVDFGETVRRSVDAVRRLRLQGRVVV.

## SEQ ID NO:26 orf2par Polypeptide

50 MATTTAKAMLERLHQYGVDFHVGFGVGREASAILFDEVEGLDFVLTRHEFTAGVMADVLRITNRPQACFATLGP  
TNLATGVATSALDRSSVIALAAQSESYDCYPNVTHQCLDSTAVMGPLTKFSVQLERGEDIVNLVDSAVLNSRIE  
GPSFISLPVDLLGAEINGTPTDAPLVRATATHALDADWRARLDEAAELVREANPLLVGSAVIRAGAVDALRALA  
ERLNI PVVTYTTAKGVLPHDHPNLYGAISGYMDGILGHPALDEIFGPADLLAIGYDYAEDLRPSMWTRGRAKTTV  
RVAVEVNPPIPELFRADIDIVTNVAEFVTAALDDATSGLAPKTRHDL SALRARVAEFADPTEYEDGMRVHQVIDCMN  
SVLDNGTFVSDIGFFRHYGVLFKSDQPYGFLTSAGCSSFGYGLPAAMAAQIARPGEPVFLIAGDGGFHSNSADIE  
TAVRLGLPIVMVVNDRNGLIELYQNLGHRSHAPAVGFGSVDFVQLAEANGCEAVRATDRTSLLAALT KGAGLG  
55 RPFLLIEVPVAYDFQSGGFAALAI